Phycologia (2001) Volume 40 (5), 440–456 Published 7 December 2001

Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae)

J.L. PAPPAS[1]*†, G.W. FOWLER[2] AND E.F. STOERMER[1]

[1]*Center for Great Lakes and Aquatic Sciences, Ann Arbor, MI 48109-1090, USA*
[2]*School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109-1115, USA*

J.L. PAPPAS, G.W. FOWLER AND E.F. STOERMER. 2001. Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae). *Phycologia* 40: 440–456.

In our study, shape descriptors were calculated for specimens from the diatom genus *Asterionella*, using the method of arc lengths and tangent angles in Fourier analysis. *Asterionella* species are character-poor, populations being distinguished mostly by subtle shape differences. Fourier analysis has been used in the past as an aid in taxonomy for many organisms. Fourier coefficients are least-squares estimators of best-fit planar closed curves that provide a quantitative measure of shape and can be useful shape descriptors in character-poor organisms. We determined the relation between number of *x,y* coordinates and the number of Fourier coefficients used in shape analysis for *Asterionella*. In a worked example, using a single extracted specimen outline, the method of calculating Fourier coefficients is demonstrated; coefficients were calculated for 100, 140, and 200 coordinates, in order to determine how many coordinates are necessary for effective shape analysis. Then, for each in a size range of *Asterionella* specimens, Fourier coefficients were calculated, using the same number of coordinates for each. Reconstructed outlines were compared graphically with the original. Statistical measures of average difference, variance, standard deviation, and coefficient of variation were calculated between *x*'s, *y*'s and Euclidean distance for the original and reconstructed outlines. From this, using 100 *x,y* coordinates, the number of Fourier coefficients necessary to give the best-fit outline over a size range from 30 to 95 μm is 22. Although we used *Asterionella* as an example, the method may be applied to any diatom valve outline.

INTRODUCTION

Morphometric analysis is especially useful for character-poor organisms. In particular, shape analysis has proved useful, because shape is empirically the initial distinguishing character of many organisms: we often use shape as a first criterion at sorting or classifying organisms, particularly those with apparently regular symmetry, such as diatoms.

Out of the many ways to accomplish shape analysis, coefficients derived from orthogonal polynomials, such as those from Fourier analysis, have proved to be suitable as shape descriptors for clams (Gevirtz 1976), ostracods (Younker & Ehrlich 1977), leaves (Kincaid & Schneider 1983), foraminifera (Lohmann 1983), mosquito wings (Rohlf & Archie 1984), and mussels (Ferson *et al.* 1985). Legendre polynomials have been used in diatom taxonomy (Stoermer & Ladewski 1982; Stoermer *et al.* 1984, 1986; Theriot & Ladewski 1986; Steinman & Ladewski 1987; Goldman *et al.* 1990), and Fourier analysis of diatom outline shape has also been attempted (Mou & Stoermer 1992). By characterizing the outline of an organism as a closed, planar curve, coefficients from orthogonal polynomials may be calculated with respect to the best fit (in a least-squares sense) of that outline. That is, shape analysis is essentially an orthogonal polynomial regression.

Few studies have been undertaken to determine the relation between the number of coordinates used and the number of Fourier coefficients calculated to obtain a best-fit outline. The Nyquist frequency has been suggested as a general guideline (e.g. Gevirtz 1976; Davis 1986). It is the harmonic of the fundamental frequency defined as $n = k/2$ (Davis 1986), where n is the number of Fourier coefficients at the Nyquist frequency, and k is the number of *x,y* coordinates used. However, the utility of the Nyquist frequency is limited because error accumulates around the periphery of an outline as more and more coefficients are added (Bennett & MacDonald 1975), and the outline may become distorted. That is, the Fourier is a waveform in which sinusoidal curves can be packed in a segment of an outline rather than a best-fit straight line, leading to overfitting of that segment (Fig. 1). In studies using Fourier methods in shape analysis, the number of harmonics is small compared to the number of coordinates used (e.g. Younker & Ehrlich 1977; Kincaid & Schneider 1983; Rohlf & Archie 1984). Each harmonic consists of at least one pair of transformed coefficients. Fewer values are used to represent an organism's outline by using Fourier shape coefficients from the harmonics.

There are many Fourier methods that can be applied in shape analysis. They are all implemented differently and one method should not be confused with another. One type of Fourier method uses radii from a centroid (e.g. Gevirtz 1976). This method is useful only when each radius intersects only one point on the curved outline (Davis 1986) (Fig. 2). Another uses the elliptical form of the Fourier based on arc lengths (e.g. Ferson *et al.* 1985), whereby the *x*-projection and *y*-projection are each represented by a pair of Fourier coefficients (Kuhl & Giardina 1982) (Fig. 3). A third method is based on arc lengths and tangent angles (e.g. Lohmann 1983), whereby change in angular direction is measured as a function of arc length (Zahn & Roskies 1972) (Fig. 4).

* Corresponding author (jlpappas@umich.edu).
† Present address: Museum of Zoology, Room 1076, 1109 Geddes Avenue, University of Michigan, Ann Arbor, MI 48109-1079. USA.

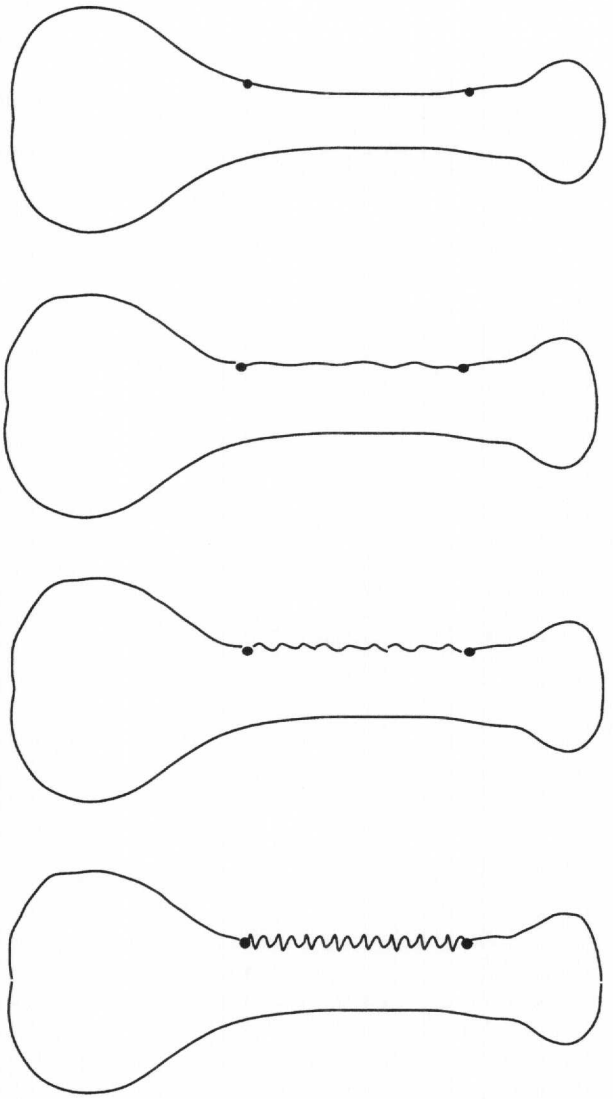

Fig. 1. Series of stylized *Asterionella* outlines illustrating overfitting between the two points. As more and more coefficients are added, this part of the curve becomes progressively more wavy or is overfitted.

In our study, diatoms from the genus *Asterionella* Hassall are used to describe how to calculate Fourier coefficients using the Fourier method based on arc lengths and tangent angles (Zahn & Roskies 1972; Bennett & MacDonald 1975; Persoon & Fu 1977). *Asterionella* was chosen because it is character-poor: shape is the outstanding feature of the valve, but this shape is particularly challenging to reproduce numerically, because there is an abrupt change from the highly curved and rounded ends to relatively straight areas along the sides. *Asterionella*'s outline is therefore not as simple as in some other diatoms, in which the outline is composed of a single geometric form (e.g. an ellipse). In addition, we investigate the relationship between number of Fourier coefficients calculated and number of *x,y* coordinates used to obtain best-fit outlines for *Asterionella* over a wide size range.

To use the method effectively to find shape similarities among all *Asterionella*, the same number of coordinates need to be assigned to each specimen, regardless of size. This ensures that pseudolandmarks can be compared from specimen

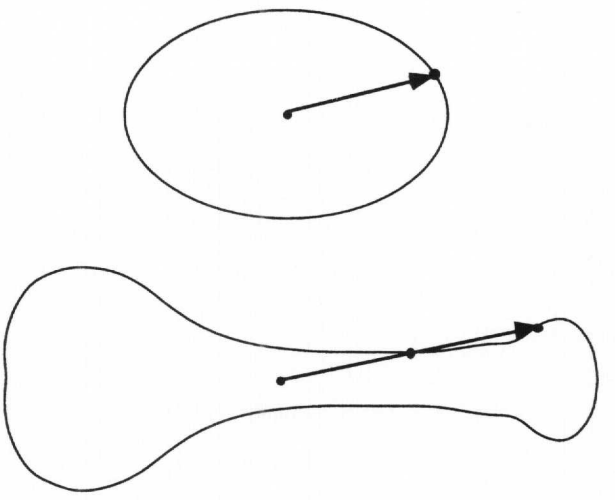

Fig. 2. Application of the Fourier method using radii drawn from a centroid: comparison of an elliptical diatom and a stylized *Asterionella*. In the elliptical diatom, each radius intersects only one point on the outline, whereas in the stylized *Asterionella*, it intersects two points.

to specimen (Lohmann 1983). This is especially important in a size diminution series: diatoms have homologous regions (e.g. capitate, rostrate, or rounded ends), but not necessarily homologous points around the outline of their valves, so that true landmark methods are not applicable. In addition, characters can be lost as a result of size diminution, so that again, landmark methods would not be appropriate. An understanding of how to use Fourier analysis based on the method of arc lengths and tangent angles will facilitate its application to problems involving shape analysis in taxonomy.

MATERIAL AND METHODS

Background on the shape analysis method

From Zahn & Roskies (1972), the normalized shape function on the interval $[0, -2\pi]$ with total arc length, L, is

$$\Phi^*(t) = \phi\left(\frac{Lt}{2\pi}\right) + t$$

where $t = 2\pi l/L$ and l is arc length. In this method, equally spaced values of t are used (Zahn & Roskies 1972; Rohlf & Archie 1984; Rohlf 1990). This shape function is invariant to translation, rotation and changes in L or dilation; therefore, the starting point on a closed curve is arbitrary (Zahn & Roskies 1972). The eigenfunction expansion gives the real-variable trigonometric Fourier series for N (the maximum number of) coefficients as the shape function

$$\Phi^*(t) = a_0 + \sum_{n}^{N} a_n\cos(nt) + b_n\sin(nt)$$

where the coefficients are

$$a_0 = -\pi - \frac{1}{L}\sum_{k}^{M} l_k\Delta\phi_k \qquad a_n = -\frac{1}{n\pi}\sum_{k}^{M}\Delta\phi_k\sin\frac{2\pi nl_k}{L} \quad \text{and}$$

$$b_n = \frac{1}{n\pi}\sum_{k}^{M}\Delta\phi_k\cos\frac{2\pi nl_k}{L}$$

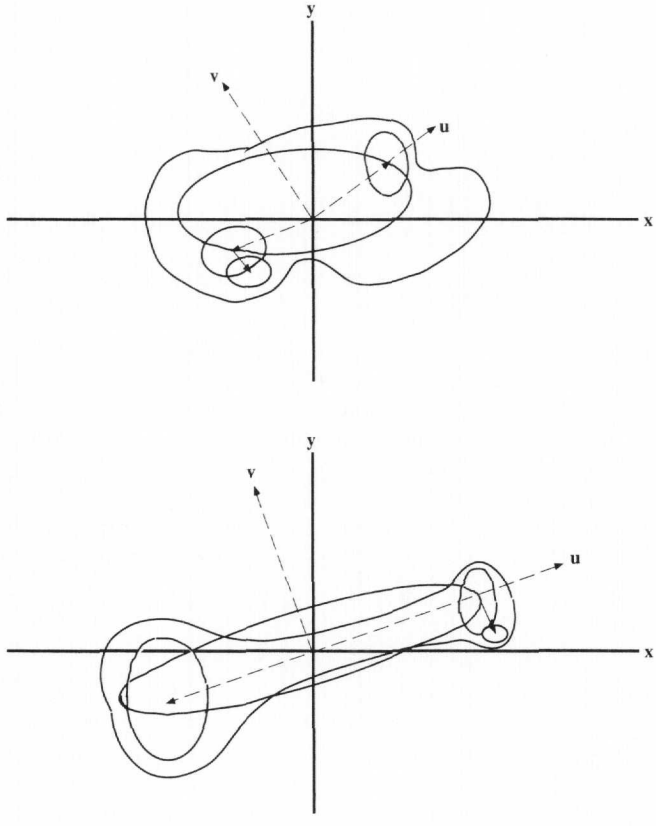

Fig. 3. Elliptical Fourier analysis: comparison of a closed curve and a stylized *Asterionella* fitted with ellipses. Vectors denote *x*- and *y*-projections for each ellipse. Two Fourier coefficients from each projection are calculated. As the number of ellipses are added to improve recovering the shape of a closed curve, more and more pairs of Fourier coefficients are calculated to represent most closely the closed curve outline.

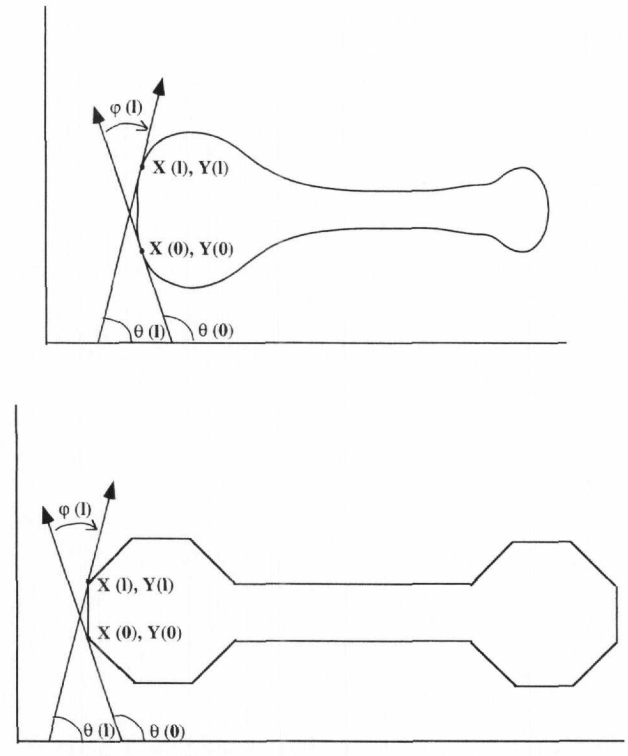

Fig. 4. Fourier analysis using the method of arc lengths and tangent angle: comparison of a stylized *Asterionella* and a polygonalized *Asterionella* depicting. Arc lengths, *l*, are calculated between successive *x,y* coordinates, and tangent angles, ϕ (with respect to arc length), are calculated relative to the angular bend in the closed curve.

where l_k is the kth arc length, $\Delta\phi_k$ is the kth angular change, n is the nth coefficient, and M is the maximum number of coordinates used (Zahn & Roskies 1972).

Converting a_n and b_n to polar coordinates results in the shape function as

$$\Phi^*(t) = a_0 + \sum_n^N A_n \cos(nt + \alpha_n),$$

where A_n is the amplitude or directed distance as

$$A_n = \sqrt{a_n^2 + b_n^2}$$

and α_n is the phase or directed angle as

$$\alpha_n = \arctan\frac{b_n}{a_n}.$$

This form of the Fourier results in the discretized version of the tangent angle function of the original shape. That is, we use the truncated polar form of the Fourier to calculate shape descriptors a_0, A_n and α_n.

Background on Fourier series

The truncated form of the Fourier is useful in shape analysis. However, this eigenfunction expansion is derived from a complex, continuous function, whose properties are of interest since they show directly how the Fourier can be used in orthogonal polynomial regression and, therefore, in shape analysis of closed curves.

The complex Fourier form of the shape function (Zahn & Roskies 1972; Persoon & Fu 1977) is

$$\Phi^*(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

where

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} \Phi^*(t) e^{-int}\, dt.$$

The complex coefficient, c_n, is related to the real coefficients by

$$c_0 = a_0 \qquad 2c_n = a_n - ib_n = A_n e^{-i\alpha_n} \quad \text{and}$$

$$2c_{-n} = a_n + ib_n = A_n e^{i\alpha_n}.$$

That is, for the real coefficients, a_n and b_n,

$$a_0 + \sum_{n=1}^{\infty} a_n \cos(nt) + b_n \sin(nt)$$

is the real part of the power series

$$a_0 + \sum_{n=1}^{\infty} (a_n - ib_n) z^n$$

on the unit circle $z = e^{int}$ (Szokefalvi-Nagy 1965). The complex conjugate of the series is

$$\sum_{n=1}^{\infty} (a_n \sin nt - b_n \cos nt)$$

To approximate the outline defined as a function, $f(t)$, we want pointwise convergence of the Fourier shape function, $\Phi^*(t)$, such that

$$\lim_{N\to\infty} \Phi^*(t) = f(t).$$

In an interval, $[g,h]$, $\Phi^*(t)$ may be determined by choosing coefficients, c_n, to minimize the difference between this function and $f(t)$. That is, in a least-squares sense, the mean square deviation is defined as

$$\lim \int_h^g [f(t) - \Phi^*(t)]^2 \rho \, dt = 0,$$

where ρdt is a positive weight function = 1.

$\Phi^*(t)$ is an orthogonal function since

$$\int_j^{j+2\pi} \sin mt \sin nt \, dt = 0,$$

$$\int_j^{j+2\pi} \cos mt \cos nt \, dt = 0,$$

$$\int_j^{j+2\pi} \sin mt \cos nt \, dt = 0 \qquad \begin{Bmatrix} (m \neq n) \\ (m = n) \end{Bmatrix}, \quad \text{and}$$

$$\int_0^{j+2\pi} e^{imt} e^{-int} \, dt = \begin{Bmatrix} 0 & (m \neq n) \\ 1 & (m = n) \end{Bmatrix}$$

where m and n are integers, $m \neq n$, and j is a real number (Edwards 1967; Kufner & Kadlec 1971). Using

$$\int_h^g [f(t) - \Phi^*(t)]^2 \rho \, dt$$

and squaring, integrating, and completing the square results in

$$\sum_{n=1}^{N} \int_h^g \Phi^*(t)^2 \rho \, dt \left\{ c_n - \frac{\int_h^g f(t)\Phi^*(t)\rho \, dt}{\int_h^g \Phi^*(t)^2 \rho \, dt} \right\}^2 + \int_h^g f(t)^2 \rho \, dt$$

$$- \sum_{n=1}^{N} \frac{\left\{ \int_h^g f(t)\Phi^*(t)\rho \, dt \right\}^2}{\int_h^g \Phi^*(t)^2 \rho \, dt}$$

The coefficients, c_n, occur only in the first sum. By making all its terms zero, minimization of a sum of squares occurs (Jackson 1941; Brown & Churchill 1993; Weinberger 1995). That is, for the real form of the Fourier, we minimize coefficients a_n and b_n to get a best-fit closed curve in a least-squares sense.

Imaging

Digital images of *Asterionella* were obtained from strewn microscope slides from two sources. One was the H. Körner collection (Botanischer Garten und Botanisches Museum Berlin-Dahlem, Berlin, Germany). Specimens from slide B Algae 25634b were used. The other source was the E.F. Stoermer collection (Center for Great Lakes and Aquatic Sciences, University of Michigan, Ann Arbor, MI 48109-1090). Specimens used were from collection (slide) 1997. Slides from Great Lakes samples were mounted in Hyrax.

A Leica DMRX compound light microscope with an oil-immersion objective, 1.40 numerical aperture, was used to image specimens at ×1000. From a Sony 3CCD model camera model 960MD and subsequently a DKC Sony 5000 camera attached to the microscope, images were transmitted to a computer, then digitized and captured using NIH Image software (version 1.62: Wayne Rasband at the US National Institutes of Health, http://rsb.info.nih.gov/nih-image/). For a complete description of the microscope and software used in image analysis, see Stoermer (1996). From the specimen outline, x,y coordinates were obtained digitally/graphically from the digitized *Asterionella* image. That is, using the imaging software, coordinates were determined using a mouse to point and click on an *Asterionella* image outline. To ensure that coordinates were equally spaced, a grid (as an image that looks like graph paper) was superimposed on the *Asterionella* image. Coordinates were checked by plotting and superimposing on the digital image to ensure that the outline represented in a pointwise fashion (Edwards 1967) was accurate for shape analysis.

Calculating arc lengths and tangent angles

To calculate Fourier coefficients, a_0, a_n, and b_n, from x,y coordinates representing the original closed curve, these coordinates need to be converted into arc lengths and tangent angles. Arc length is

$$l_k = \sqrt{\Delta x^2 + \Delta y^2}$$

where l_k is the kth arc length. Cumulative arc length (Zahn & Roskies 1972) is

$$\sum_{k=1}^{M} l_k$$

from the start to the kth point on the curve. To obtain the closed curve, the first set of x,y coordinates is also used as the last set for the last cumulative arc length. Tangent angles, as angular bend, ϕ, in a clockwise (negative) direction, are initially calculated by

$$\phi = -\arctan\left(\frac{\Delta y}{\Delta x}\right).$$

At each point on a closed curve, we want to calculate change in angular bend, $\Delta\phi$, where $\phi(l)$ is the cumulative angular bend function (with respect to arc length), or net angular bend from the starting point to a point on the curve, l (arc length), on a closed curve. Total net angular bend is $\phi(L) = -2\pi$. From this, $\phi(l)$ will need to be normalized on the interval $[0, -2\pi]$ to $\Phi^*(t)$, to be used in the Fourier expansion of the shape function. That is, −arctangent values have to be corrected for direction and magnitude on the interval $[0, -2\pi]$ before Fourier coefficients can be calculated.

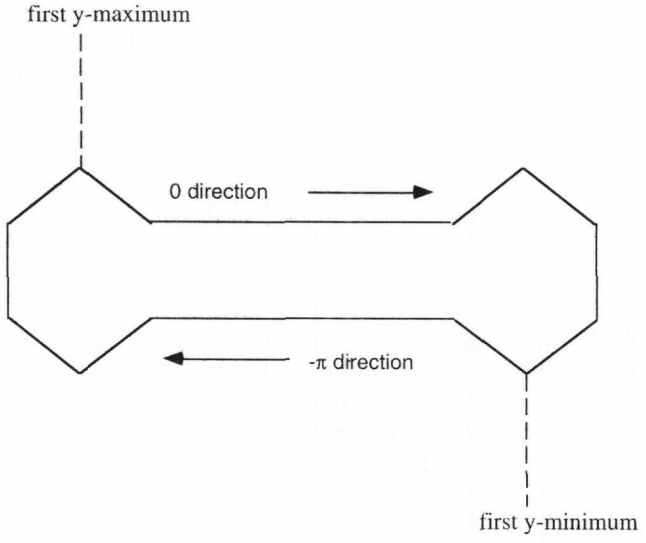

Fig. 5. Polygonalized *Asterionella*, divided into top and bottom halves. The first *y*-maximum and the first *y*-minimum are the points that define the halves. The top half is the zero sector; the bottom half is the $-\pi$ sector.

Direction of angular bend and correction of −arctangent values

We will use stylized polygonalized drawings of *Asterionella*'s shape to describe the process of correcting −arctangent values. With respect to *x,y* coordinates, the outline of *Asterionella* is divided into halves, where one half is defined to begin where the *y*-value is a first maximum and the *x*-value is small, but subsequent *x*-values are increasing (Fig. 5). The top half ends where the first *y*-value minimum occurs and the *x*-value is large, but subsequent *x*-values are decreasing (Fig. 5). With respect to dividing the outline of *Asterionella* into 0 (or -2π) and π sectors, the top half of *Asterionella* is in the zero (or -2π) sector or direction and the bottom half of *Asterionella* is in the $-\pi$ sector or direction (Fig. 5). After calculating negative arctangent (−arctangent) values, the next step is to add $-\pi$ to the −arctangent values for the bottom half of the outline.

Direction of angular bend follows a general pattern (Fig. 6). Following the outline in a clockwise (negative) direction, whether the curve bends outward or inward for the top or bottom half determines the sign of the value added to the −arctangent values (Fig. 6). In general, the way to correct for direction involves using $\pi/2$ fractions (Bennett & MacDonald 1975): a value of $\leq \pi/2$ needs to be added or subtracted from the angular bend. In the top half of the outline, where the angle first increases, $\pi/2$ or less would be added. Likewise, in the bottom half of the outline, where the angle first decreases, $\pi/2$ or less would be added. Complex three-sided outlines can be analysed based on $\pi/4$ fractions. However, diatom shapes which have much less curvature, approaching circular to oval or triangular shapes, can be more simply analysed using the Fourier method based on centroids and radii. For our purposes and using *Asterionella* as a worked example, guidelines based on multiples of $\pi/2$ fractions are specifically covered in the next section.

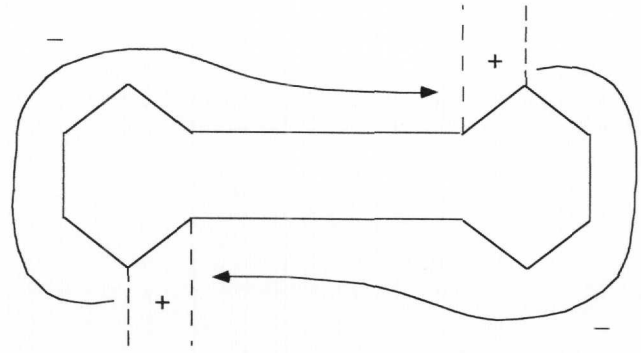

Fig. 6. Polygonalized *Asterionella*, showing the positive and negative sectors of the curve where initial $-\pi/2$ fractions are added or subtracted.

Magnitude of angular bend

In general, change in angular bend means recovering all the peaks and valleys around the periphery of an outline. A circle, the simplest closed curve, is 2π radians. A triangle is the simplest figure with edges, and diatoms that are basically triangular may be corrected by starting with this figure. A square is the next figure in this series, with four edges (Fig. 7). In any such angular figure, there is a change in angular bend at

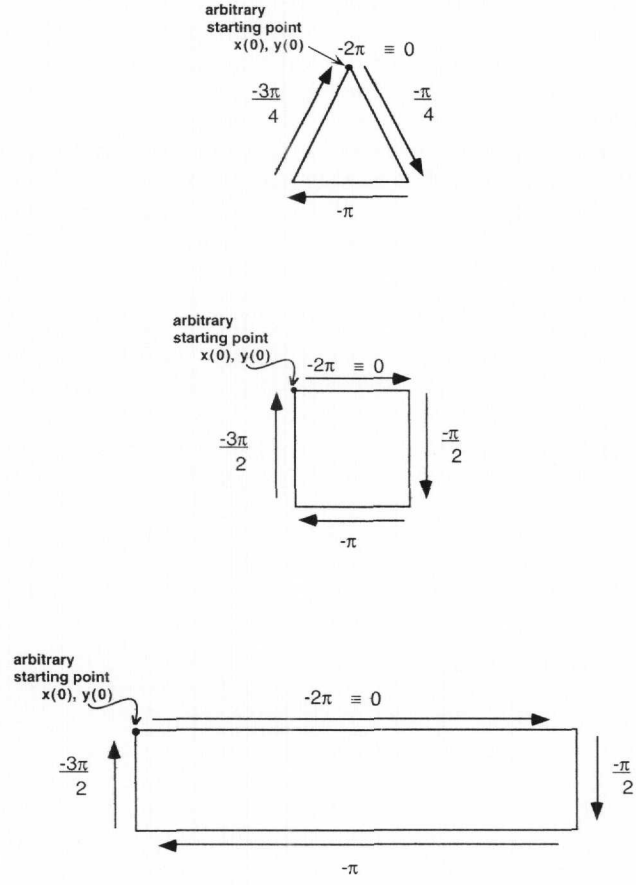

Fig. 7. Series of the simplest polygonal closed curves, a triangle, a square, and a rectangle, with arbitrary starting point x_0, y_0. Magnitude of angular bend is in $-\pi/4$ increments around the periphery of the triangle and $-\pi/2$ increments around the peripheries of the square and rectangle.

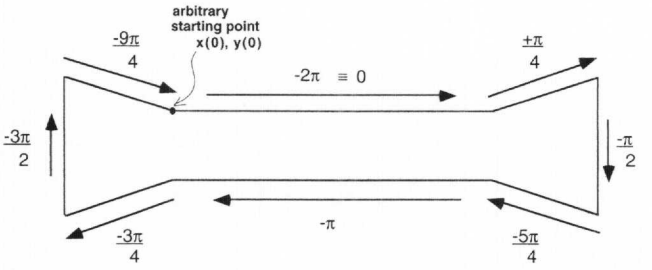

Fig. 8. Concavity illustrated for angular bends around the periphery of a closed curve.

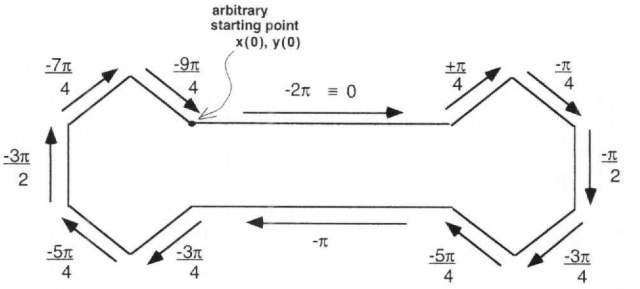

Fig. 10. Polygonalized *Asterionella* with a combination of concave and convex angular bends around the periphery of the curve.

the vertex between two edges (e.g. Fig. 7). *Asterionella* is obviously more complex, and to illustrate recovering peaks and valleys in the outline of *Asterionella,* we use a stylized (polygonalized) drawing, in which *Asterionella* has been given straight edges and made into a crude polygon. The series of polygons shown in Figs 5–10 can be regarded as progressive approximations to the shape of *Asterionella.* As the edges illustrating angular bend become smaller and smaller, the polygonized *Asterionella*'s outline becomes smoother and smoother, coming more and more to resemble the actual diatom valve outline, rather than a stylized drawing.

To determine the magnitude of angular bend around the periphery of the outline, a value of $\leq \pi/2$ is added to or subtracted from the −arctangent values. This is accomplished using a series of polygonal curves that approach an *Asterionella* outline (Figs 7–10). These polygonal curves serve as guidelines to proceed from a polygon with distinct angles to a polygon with approximately smooth curvature. Since we are dealing with a pennate diatom outline, the simplest polygonal figure from which to start is a rectangle (Fig. 7, third drawing). An arbitrary starting point is chosen, such as x_0,y_0 on the top half of the outline (Fig. 7, third drawing). Any other point on the closed curve may be chosen as a starting point (additional information on choosing a starting point is given in the Discussion). In a clockwise direction, the uppermost horizontal edge is equal to zero, a 90° bend equals $-\pi/2$, a subsequent 90° bend equals $-\pi$, another 90° bend equals $-3\pi/2$, and the final 90° bend to the starting point ends at -2π.

The next two polygons in the series add angular bends to produce either concavity (Fig. 8) or convexity (Fig. 9) and this reflects the general pattern of angular bend. Now the angular bends are subdivided and defined in multiples of $-\pi/4$. By combining the angular bends from these polygons into a shape more closely resembling *Asterionella,* finer and finer angular bends are defined (Figs 10, 11). Concavity or convexity can occur at any point on the outline. To determine the magnitude of angular bend, addition or subtraction of $-\pi/4$ or a multiple of $-\pi/4$ is performed, depending on which half of *Asterionella* is being considered. The top half can be $\pm\pi/4$, and the bottom half can be a multiple of $\pm\pi/4$, since this is in the $-\pi$ sector (Figs 5, 6, 10, 11).

Further subdivisions of $-\pi/4$, then $-\pi/8$ or higher multiples, produce smoother and smoother curvature from vertices of angular bends. To determine the exact value to add or subtract, −arctangent values are compared to the interval values in radians given in Table 1 (equivalent values are given also in degrees). The maximum interval defined is based on angular bend equal to $-\pi/2$, then halved for preceding intervals until the minimum value approaches zero (Table 1). Adding or subtracting the appropriate value once to each −arctangent value based on interval values (Table 1) will produce the general shape. For complex forms, getting a curve fit that is closest to the original shape may require additional manipulation of −arctangent values.

Figure 12 illustrates a composite of Figs 5, 6, and 8–11. The total number of radians at each angular bend must result in a value within a specific range (Fig. 12). Designate an arbitrary starting point on the top half. From this, the total number of radians at each angular bend cannot exceed $-9\pi/4$ at the first decline, $+\pi/4$ at the first incline, $-\pi/2$ at the next decline, and $-3\pi/4$ at the decline after $\phi = -\pi/2$ (Fig. 12). For the bottom half, the total number of radians at each angular bend cannot exceed $-5\pi/4$ at the first incline, $-3\pi/4$ at the first decline, $-3\pi/2$ at the next incline, and $-7\pi/4$ at the incline after $\phi = -3\pi/2$ (Fig. 12). At a given point on the most highly curved part of the outline, the total number of radians might be in a specific range somewhat greater or less than the range defined just prior to this point and will need to be adjusted.

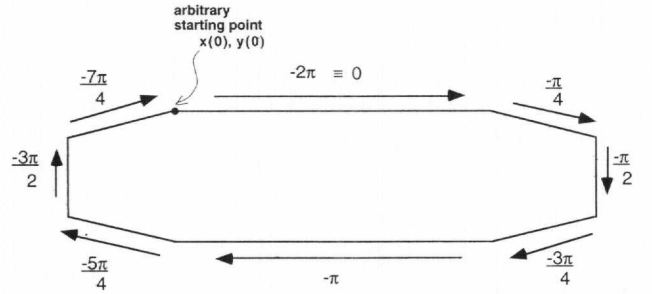

Fig. 9. Convexity illustrated for angular bends around the periphery of a closed curve.

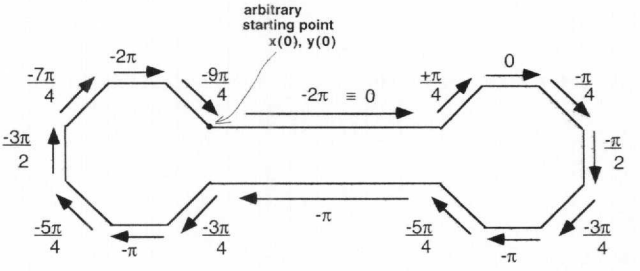

Fig. 11. A more refined polygonalized model of *Asterionella,* with an increased number of angular bends relative to Fig. 10.

Table 1. Correction of −arctangent values for angular bend based on fraction of π/2.

If −arctangent is in interval (degrees)	If −arctangent is in interval (radians)	Add or subtract π/2 fraction
> 0–1.40625	> 0–0.024544	π/128
1.40626–2.8125	0.024525–0.049087	π/64
2.8126–5.625	0.049087–0.098175	π/32
5.626–11.25	0.098176–0.19635	π/16
11.26–22.5	0.19636–0.3927	π/8
22.6–45.0	0.3928–0.7854	π/4
45.1–90.0	0.7855–1.5708	π/2

In addition, at the first *y*-maximum and the first *y*-minimum, the $-\pi/2$ fraction added is greater than the first fraction added. That is, the total number of radians at the first *y*-maximum is a value close to but not exceeding -2π. At the first *y*-minimum, the total number of radians is a value close to but not exceeding $-\pi$. This ensures recovery of the angular bend at these points.

After calculating angular bends, the change in angular bend, $\Delta\phi$, is calculated. The last change in angle is calculated between the last and first points. Since the curve is closed and the first point is used twice, -2π is added to correct the last change in angular bend.

Calculating amplitudes and phase angles

From the arc lengths and corrected tangent angles, polar Fourier coefficients, amplitudes and phase angles (as stated previously), are calculated as

$$A_n = \sqrt{a_n^2 + b_n^2}$$

and

$$\alpha_n = \arctan\frac{b_n}{a_n},$$

respectively. Unlike the amplitudes, phase angles are not invariant. Phase angles are modulo 2π, except for the *x*-direction of the outline, which is modulo π (Zahn & Roskies 1972). This means that phase angles can be multiples of 2π (or π for the *x*-direction) and still be correct as Fourier coefficients, whereas amplitudes do not change in value. For example, a phase angle at 0° is the same as one at 360° (2π), which is the same as one at 720° (4π) and so on. After calculating initial values for the phase angles, π is added to those that have a corresponding Fourier coefficient for the (*x*-direction) cosine term, a_n, that is negative. For negative values of phase angles that still result after this, 2π is added. The final result is the corrected positive phase angles. From this, all 2π multiples of the phase angles would be correct Fourier coefficients.

Reconstructing an outline from Fourier coefficients

Cartesian or *x,y* coordinates may be calculated from Fourier coefficients in the following way. Define a point on the reconstructed outline as

$$Z(l) = \langle x(l), y(l)\rangle = Z(0) + \int_0^l e^{i\theta(l)}\, dl$$

with

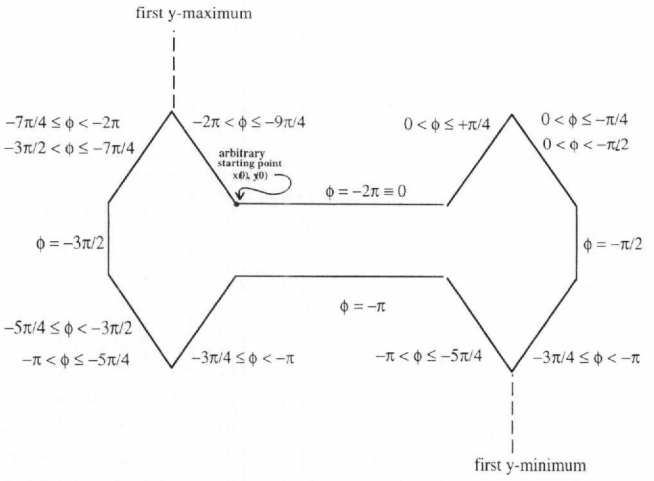

Fig. 12. Polygonalized *Asterionella* [cf. Fig. 10, with ranges defining the total number of radians at angular bends around the periphery of the curve (magnitude)]. Not all of the possible angular bends approaching smoothness of the closed curve are shown; therefore, ranges of radians are given. The first *y*-maximum, first *y*-minimum, and the arbitrary starting point are designated.

$$x(l) = x(0) + \int_0^l \cos\,\theta(l)\, dl \qquad y(l) = y(0) + \int_0^l \sin\,\theta(l)\, dl$$

where

$$\theta(l) = -t + \delta + a_0 + \sum_{n=1}^{N} A_n\cos(nt - \alpha_n)$$

That is,

$$Z(l) = Z(0) + \frac{L}{2\pi}\int_0^{2\pi l/L} \exp\left\{i\left[-t + \delta + a_0 + \sum_{n=1}^{N} A_n\cos(nt - \alpha_n)\right]\right\} dt$$

where δ is the initial angular bend, ϕ_0, at starting point $Z(0)$ (Zahn & Roskies 1972). Initial angular bend is equal to zero if two conditions are met. First, coordinates must be obtained when the specimen outline is aligned with the *x*-axis. Second, the change in angular bend at the first arc length must be zero. If the first condition is met but not the second, the initial angular bend is equal to the angular bend at the first arc length. If the first condition is not met but the second one is, the initial angular bend would need to be corrected by number of radians rotated from zero degrees. If neither condition is met, the initial angular bend is equal to the angular bend at the first arc length plus the number of radians rotated from zero degrees.

The relationship between the number of coordinates used and the number of coefficients needed to reconstruct a best-fit curve was determined. One hundred, 140, or 200 equally spaced *x,y* coordinates were designated around the periphery of the outline of an *Asterionella* valve from slide 25634b (Körner collection). Then *Asterionella* specimens of different sizes from slides 25634b and 1997 (Stoermer collection) were used to determine whether a best-fit outline could be reconstructed for all of them, regardless of size, using the same number of coefficients.

To determine best fit, the original coordinates were compared with the reconstructed outline coordinates by calculating

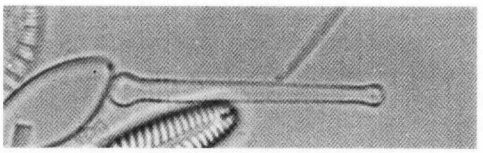

Fig. 13. Digitized image of *Asterionella* specimen 25634b-8. The outline from this specimen was used to illustrate how to perform shape analysis calculating Fourier coefficients.

average difference, variance, standard deviation, and coefficient of variation for x-coordinates, y-coordinates, and Euclidean distance between the closed curves, on a pointwise basis. The smallest difference between the original and reconstructed outlines on a pointwise basis for the majority of the measures indicated the number of coefficients to use for a best-fit outline. Reconstructed and original outlines were graphed in a pointwise fashion to show how well the reconstructed outline reproduced the original from coordinate to coordinate.

RESULTS

To illustrate the method, 100 x,y coordinates were used from the outline of *Asterionella* specimen 25634b-8 (Fig. 13). Tangent angles are presented in Table 2. The second column gives the initial $-$arctangents at each x,y coordinate. The third column shows that the bottom half of the outline has $-\pi$ added to the $-$arctangent values (see Fig. 11). The fourth column shows the $\pi/2$ fraction added or subtracted to the $-$arctangent value, based on Table 1. The sixth column shows the $\pi/2$ fraction or more added or subtracted to achieve final $-$arctangent values, based on Fig. 10. The seventh column reiterates the ranges from Fig. 12. The final $-$arctangent values of angular bend are given in the last column.

In Table 3, amplitudes are given in the second column. In addition, phase angles and their final corrected positive values are given. In the third column, initial phase angles are given. In the fifth column, phase angle values are corrected by the addition of π if the cosine term's coefficient was negative (Table 3, fourth column); that is, this correction is mod π. In the seventh column, phase angles that were still negative have 2π added to them; that is, phase angles are mod 2π. The final phase angles are given in the last column (Table 3).

An example of the reconstructed outlines using 0–30 Fourier coefficients for this *Asterionella* specimen is shown in Fig. 14. Only even coefficients are used, since odd coefficients are indicators of symmetry, not shape, of the closed curve. The simplest reconstruction is an ellipse (Fig. 14, first reconstructed outline). As more coefficients are added, the outline approaches that of the original coordinates (Fig. 14): the ellipse undergoes deformation as coefficients are added to achieve the desired outline. How hard one induces deformation at equally spaced points on a circle/oval is determined by the magnitude of amplitudes (Zahn & Roskies 1972).

The relationship between the number of coordinates designated and the number of Fourier coefficients calculated was investigated in relation to the effectiveness of shape description. For example, the outline from *Asterionella* specimen 25634b-8 (Fig. 13), which is 30.2 μm long, was represented by 100, 140, or 200 x,y coordinates. Initially, x,y coordinates were compared to the originals empirically by graphing. For 100 x,y coordinates along the original outline, reconstructed approximate best-fit outlines were obtained from 18, 20, 22, or 24 Fourier coefficients (Fig. 15). For 140 x,y coordinates, reconstructed approximate best-fit outlines were obtained using 16, 18, 20, or 22 Fourier coefficients (Fig. 16), whereas for 200 x,y coordinates, the number was 14, 16, 18, or 20 (Fig. 17). Beyond 24 coefficients for 100 coordinates, 22 coefficients for 140 coordinates, and 20 coefficients for 200 coordinates, overfitted closed curves were produced.

When the specimen outline was represented using 100 coordinates, the smallest average difference between x coordinates was at 22 coefficients. The smallest average difference between y coordinates was at 18 coefficients. The smallest average distance between x,y coordinates was at 22 coefficients. For x coordinates, 22 coefficients produced the smallest variance. For y coordinates, 24 coefficients produced the smallest variance. For distance between x,y coordinates, 18 coefficients produced the smallest variance. The smallest difference in coefficient of variation for x coordinates was at 18 coefficients. However, for y coordinates, the smallest difference in coefficient of variation was at 22 coefficients. The smallest distance between x,y coordinate distances for coefficient of variation was at 24 coefficients (Table 4).

For this specimen outline using 140 coordinates, the smallest average difference between x coordinates and between y coordinates was at 22 coefficients. The smallest average distance between x,y coordinate distances was produced at 18 coefficients. Variances were smallest at 22 coefficients for x coordinates and y coordinates and at 16 coefficients for the distance between x,y coordinates. For coefficients of variation, 20, 18, and 16 Fourier coefficients produced the smallest differences for x coordinates, y coordinates, and x,y coordinate distances, respectively (Table 5).

Using 200 coordinates, the smallest average difference between x coordinates was at 20 coefficients. For y coordinates and x,y coordinate distances, 20 and 16 coefficients produced the smallest average differences, respectively. The smallest variances were at 18, 16, and 20 coefficients for x coordinates, y coordinates, and x,y coordinate distances, respectively. The smallest differences for coefficients of variation were 14, 14, and 18 coefficients for x coordinates, y coordinates, and x,y coordinate distances, respectively (Table 6).

Along with the 30.2 μm specimen, the smallest to be found in the collections used, *Asterionella* outlines were obtained from a range of other specimens, the longest being 94.6 μm. The fewest number of coordinates in this analysis, 100, was used to cover all size ranges of *Asterionella,* because computation time is slightly increased as more coordinates are used. For illustration, reconstructed outlines were produced from 18, 20, 22, and 24 coefficients for three specimens covering the size range of *Asterionella,* from the smallest specimen to 51.9 μm (Fig. 18) to 94.6 μm (Fig. 19).

For the 51.9 μm specimen, 25634b-10, the smallest average difference for x coordinates was at 18 coefficients, whereas for y coordinates and x,y coordinate distances it was at 20 coefficients. The smallest variances for x coordinates, y coordinates and x,y coordinate distances were at 18, 22, and 22 coefficients, respectively. The smallest coefficients of variation for x coordinates, y coordinates, and x,y coordinate dis-

Table 2. Values and corrections for −arctangents of *Asterionella* 25634b-8 outline. See Table 1 for intervals of $-\pi/2$ fractions and Fig. 12 for ranges in radians of angular bends.

Number of *x,y* coordinate	−arctangent value	Addition of $-\pi$ to bottom half	$\pi/2$ fraction to be added or subtracted according to Table 1	Number of radians subtotal	Addition or subtraction of another $\pi/2$ fraction or more according to Fig. 12	Equivalent angle or ranges from Fig. 12	Final −arctangent value
1	0	0		0		0	0
2	0	0		0		0	0
3	0	0		0		0	0
4	0	0		0		0	0
5	0	0		0		0	0
6	0	0		0		0	0
7	0	0		0		0	0
8	0	0		0		0	0
9	0	0		0		0	0
10	0	0		0		0	0
11	0	0		0		0	0
12	0	0		0		0	0
13	0	0		0		0	0
14	0	0		0		0	0
15	0	0		0		0	0
16	0	0		0		0	0
17	0	0		0		0	0
18	0	0		0		0	0
19	0	0		0		0	0
20	0	0		0		0	0
21	0	0		0		0	0
22	0	0		0		0	0
23	0	0		0		0	0
24	0	0		0		0	0
25	0	0		0		0	0
26	0	0		0		0	0
27	0.099669	0	$-\pi/16$	−0.096681		$0 < \phi \leq -\pi/4$	−0.096681
28	0.099669	0	$-\pi/16$	−0.096681		$0 < \phi \leq -\pi/4$	−0.096681
29	0.099669	0	$-\pi/16$	−0.096681		$0 < \phi \leq -\pi/4$	−0.096681
30	0	0		0		0	0
31	0	0		0		0	0
32	0	0		0		0	0
33	0	0		0		0	0
34	0	0		0		0	0
35	0.099669	0	$-\pi/16$	−0.096681		$0 < \phi \leq -\pi/4$	−0.096681
36	−0.1974	0	$+\pi/8$	0.1953		$0 < \phi \leq +\pi/4$	0.1953
37	−0.099669	0	$+\pi/16$	0.096681		$0 < \phi \leq +\pi/4$	0.096681
38	0.69474	0	$-\pi/4$	−0.09066	$-\pi/4$	$0 < \phi < -\pi/2$	−0.87606
39	0.7854	0	$-\pi/4$	0	$-\pi/4$	$0 < \phi \leq -\pi/4$	−0.7854
40	1.5708	0	$-\pi/2$	0	$-\pi/2$	$-\pi/2$	−1.5708
41	−0.87606	0	$-\pi/2$	−2.44690		$-3\pi/4 < \phi \leq -\pi$	−2.4469
42	−0.46365	0	$-\pi/4$	−1.24905	$-\pi/2$	$-3\pi/4 < \phi \leq -\pi$	−2.8198
43	0.099669	$-\pi$	$-\pi/16$	−3.2383		$-\pi < \phi \leq -5\pi/4$	−3.2383
44	0.099669	$-\pi$	$-\pi/16$	−3.2383		$-\pi < \phi \leq -5\pi/4$	−3.2383
45	0.1974	$-\pi$	$-\pi/8$	−3.3369		$-\pi < \phi \leq -5\pi/4$	−3.3369
46	0	$-\pi$		−3.1416		$-\pi$	−3.1416
47	0	$-\pi$		−3.1416		$-\pi$	−3.1416
48	0	$-\pi$		−3.1416		$-\pi$	−3.1416
49	0	$-\pi$		−3.1416		$-\pi$	−3.1416
50	0	$-\pi$		−3.1416		$-\pi$	−3.1416
51	0	$-\pi$		−3.1416		$-\pi$	−3.1416
52	0	$-\pi$		−3.1416		$-\pi$	−3.1416
53	0	$-\pi$		−3.1416		$-\pi$	−3.1416
54	0	$-\pi$		−3.1416		$-\pi$	−3.1416
55	0	$-\pi$		−3.1416		$-\pi$	−3.1416
56	0	$-\pi$		−3.1416		$-\pi$	−3.1416
57	0	$-\pi$		−3.1416		$-\pi$	−3.1416
58	0	$-\pi$		−3.1416		$-\pi$	−3.1416
59	0	$-\pi$		−3.1416		$-\pi$	−3.1416
60	0	$-\pi$		−3.1416		$-\pi$	−3.1416
61	0	$-\pi$		−3.1416		$-\pi$	−3.1416
62	0	$-\pi$		−3.1416		$-\pi$	−3.1416
63	0	$-\pi$		−3.1416		$-\pi$	−3.1416

Table 2. Continued.

Number of *x,y* coordinate	−arctangent value	Addition of $-\pi$ to bottom half	$\pi/2$ fraction to be added or subtracted according to Table 1	Number of radians subtotal	Addition or subtraction of another $\pi/2$ fraction or more according to Fig. 12	Equivalent angle or ranges from Fig. 12	Final −arctangent value
64	0	$-\pi$		−3.1416		$-\pi$	−3.1416
65	0	$-\pi$		−3.1416		$-\pi$	−3.1416
66	0	$-\pi$		−3.1416		$-\pi$	−3.1416
67	0	$-\pi$		−3.1416		$-\pi$	−3.1416
68	0	$-\pi$		−3.1416		$-\pi$	−3.1416
69	0	$-\pi$		−3.1416		$-\pi$	−3.1416
70	0	$-\pi$		−3.1416		$-\pi$	−3.1416
71	0	$-\pi$		−3.1416		$-\pi$	−3.1416
72	0	$-\pi$		−3.1416		$-\pi$	−3.1416
73	0	$-\pi$		−3.1416		$-\pi$	−3.1416
74	−0.099669	$-\pi$	$+\pi/16$	−3.0449		$-3\pi/4 < \phi \leq -\pi$	−3.0449
75	−0.099669	$-\pi$	$+\pi/16$	−3.0449		$-3\pi/4 < \phi \leq -\pi$	−3.0449
76	−0.049958	$-\pi$	$+\pi/32$	−3.0934		$-3\pi/4 < \phi \leq -\pi$	−3.0934
77	0.049958	$-\pi$	$-\pi/32$	−3.1898		$-\pi < \phi \leq -5\pi/4$	−3.1898
78	0.099669	$-\pi$	$-\pi/16$	−3.2383		$-\pi < \phi \leq -5\pi/4$	−3.2383
79	0.099669	$-\pi$	$-\pi/16$	−3.2383		$-\pi < \phi \leq -5\pi/4$	−3.2383
80	−0.049958	$-\pi$	$+\pi/32$	−3.0934		$-3\pi/4 < \phi \leq -\pi$	−3.0934
81	−0.14889	$-\pi$	$+\pi/16$	−3.0941		$-3\pi/4 < \phi \leq -\pi$	−3.0941
82	−0.099669	$-\pi$	$+\pi/16$	−3.0449		$-3\pi/4 < \phi \leq -\pi$	−3.0449
83	−0.1974	$-\pi$	$+\pi/8$	−2.9463	$+\pi/8$	$-3\pi/4 < \phi \leq -\pi$	−2.5536
84	−0.1974	$-\pi$	$+\pi/8$	−2.9463	$+\pi/8$	$-3\pi/4 < \phi \leq -\pi$	−2.5536
85	0.1974	$-\pi$	$-\pi/8$	−3.3369	$-\pi/8$	$-\pi < \phi \leq -5\pi/4$	−3.7296
86	0.1974	$-\pi$	$-\pi/8$	−3.3369	$-\pi/8$	$-\pi < \phi \leq -5\pi/4$	−3.7296
87	0.1974	$-\pi$	$-\pi/8$	−3.3369	$-\pi/8$	$-\pi < \phi \leq -5\pi/4$	−3.7296
88	0.54042	$-\pi$	$-\pi/4$	−3.3866	$-\pi/4$	$-\pi < \phi \leq -5\pi/4$	−4.172
89	1.2278	$-\pi$	$-\pi/2$	−3.4846		$-\pi < \phi \leq -5\pi/4$	−3.4846
90	1.1071	$-\pi$	$-\pi/2$	−3.6052		$-\pi < \phi \leq -5\pi/4$	−3.6052
91	−0.87606	$-\pi$	$-\pi/2$	−5.5885	$-\pi/4$	$-2\pi < \phi \leq -9\pi/4$	−6.3738
92	−1.1071	$-\pi$	$-\pi/2$	−5.8195	$-\pi/4$	$-2\pi < \phi \leq -9\pi/4$	−6.6049
93	−0.7854	$-\pi$	$-\pi/4$	−3.1416	$-\pi/4$	$-3\pi/2 < \phi \leq -7\pi/4$	−5.4978
94	0.67474	$-\pi$	$-\pi/4$	−4.6017	$-\pi/4$	$-3\pi/2 < \phi \leq -7\pi/4$	−5.3871
95	−0.099669	$-\pi$	$-\pi/16$	−3.4376	$-3\pi/4$	$-7\pi/4 \leq \phi < -2\pi$	−5.7938
96	0.099669	-2π (0)	$-\pi/16$	−6.3799		$-2\pi < \phi \leq -9\pi/4$	−6.3799
97	0.099669	-2π (0)	$-\pi/16$	−6.3799		$-2\pi < \phi \leq -9\pi/4$	−6.3799
98	0.29146	-2π (0)	$-\pi/8$	−0.1012	$-\pi/8$	$-2\pi < \phi \leq -9\pi/4$	−6.7771
99	0.099669	-2π (0)	$-\pi/16$	−6.3799		$-2\pi < \phi \leq -9\pi/4$	−6.3799
100	0.38051	-2π (0)	$-\pi/8$	−0.0122	$-\pi/8$	$-2\pi < \phi \leq -9\pi/4$	−6.6881

tances were at 20, 24, and 22 coefficients, respectively (Table 7).

For the 94.6 μm specimen, 1997-1, the smallest average difference for *x* coordinates was at 24 coefficients. For *y* coordinates and *x,y* coordinate distances, however, the smallest average difference was at 22 coefficients The smallest variances for *x* coordinates, *y* coordinates, and *x,y* coordinate distances were produced at 24, 20, and 22 coefficients. The coefficient of variation was smallest at 20 coefficients for both *x* and *y* coordinates and at 18 coefficients for *x,y* coordinate distances (Table 8).

DISCUSSION

This method of shape analysis requires careful thought and patience and is not initially the easiest method to use. Determining the magnitude of corrections to the −arctangent values for angular bend is complicated, but important. In theory, an infinite number of peaks and valleys can be formed on a closed curve. However, once the method is learned, application becomes easier, and a reliable way to quantify an important taxonomic character, namely shape, can be used.

The method, as described here, can be applied to any closed curve, and so any diatom valve shape can be analysed, not just those in which each radius from the centroid intersects the outline only once (as required for the method of Gevirtz 1976). Table 1 provides general guidelines for correcting tangent angles for any closed curve. By halving $\pi/2$ an indefinite number of times, with multiples thereof, any change in angular bend can be represented mathematically. This is concomitant with the idea that a circle can be deformed in an infinite number of ways.

We recommend using shape analysis on flat, whole valves, but our method is also applicable to partial outlines, and it will be interesting to apply the method to partial outlines, once a suitable database of whole valve shape coefficients has been amassed. As noted in the Methods section, for whole valves the first *x,y* coordinate is also used as the last one. If the first is not used in this way to create a closed form, a partial outline

Table 3. Amplitudes and phase angle calculations. First, phase angles are corrected by $+\pi$ if cosine term coefficients are negative. If phase angles are still negative, they are corrected by $+2\pi$.

Harmonic	Amplitudes	Initial phase angles	Cosine term coefficient,*a*	Correction with $+\pi$	Result of first correction	Correction with 2π	Final phase angles
1	0.013818	0.438087	−0.012513	$\alpha 1 + \pi$	3.579679	$\alpha 1 + 0$	3.579679
2	1.003617	0.522054	−0.869932	$\alpha 2 + \pi$	3.663647	$\alpha 2 + 0$	3.663647
3	0.018585	−0.651801	−0.014775	$\alpha 3 + \pi$	2.489792	$\alpha 3 + 0$	2.489792
4	0.499812	−0.484703	−0.442240	$\alpha 4 + \pi$	2.656889	$\alpha 4 + 0$	2.656889
5	0.006486	−1.077410	0.003072	$\alpha 5 + 0$	−1.077410	$\alpha 5 + 2\pi$	5.205776
6	0.318743	−1.468506	−0.032547	$\alpha 6 + \pi$	1.673086	$\alpha 6 + 0$	1.673086
7	0.026749	0.525497	−0.023140	$\alpha 7 + \pi$	3.667090	$\alpha 7 + 0$	3.667090
8	0.206491	0.646142	0.164865	$\alpha 8 + 0$	0.646142	$\alpha 8 + 0$	0.646142
9	0.030049	0.030594	−0.030035	$\alpha 9 + \pi$	3.172187	$\alpha 9 + 0$	3.172187
10	0.119536	−0.461638	0.107023	$\alpha 10 + 0$	−0.461638	$\alpha 10 + 2\pi$	5.821547
11	0.052132	−0.451236	−0.046914	$\alpha 11 + \pi$	2.690357	$\alpha 11 + 0$	2.690357
12	0.062471	1.524299	−0.002904	$\alpha 12 + \pi$	4.665892	$\alpha 12 + 0$	4.665892
13	0.062208	−1.514796	−0.003482	$\alpha 13 + \pi$	1.626797	$\alpha 13 + 0$	1.626797
14	0.039871	0.423990	−0.036341	$\alpha 14 + \pi$	3.565583	$\alpha 14 + 0$	3.565583
15	0.037805	0.422178	0.034486	$\alpha 15 + 0$	0.422178	$\alpha 15 + 0$	0.422178
16	0.036136	−0.637287	−0.029043	$\alpha 16 + \pi$	2.504305	$\alpha 16 + 0$	2.504305
17	0.003506	−0.502721	0.003072	$\alpha 17 + 0$	−0.502721	$\alpha 17 + 2\pi$	5.780464
18	0.031746	1.496898	0.002344	$\alpha 18 + 0$	1.496898	$\alpha 18 + 0$	1.496898
19	0.014392	1.163087	0.005707	$\alpha 19 + 0$	1.163087	$\alpha 19 + 0$	1.163087
20	0.027484	0.708128	0.020876	$\alpha 20 + 0$	0.708128	$\alpha 20 + 0$	0.708128
21	0.020964	0.386867	0.019415	$\alpha 21 + 0$	0.386867	$\alpha 21 + 0$	0.386867
22	0.032960	−0.139757	0.032639	$\alpha 22 + 0$	−0.139757	$\alpha 22 + 2\pi$	6.143428
23	0.034668	−0.594135	0.028727	$\alpha 23 + 0$	−0.594135	$\alpha 23 + 2\pi$	5.689051
24	0.042972	−1.199261	0.015601	$\alpha 24 + 0$	−1.199261	$\alpha 24 + 2\pi$	5.083925
25	0.051739	1.347617	−0.011451	$\alpha 25 + \pi$	4.489210	$\alpha 25 + 0$	4.489210
26	0.053033	0.930982	−0.031663	$\alpha 26 + \pi$	4.072575	$\alpha 26 + 0$	4.072575
27	0.063675	0.206378	−0.062324	$\alpha 27 + \pi$	3.347971	$\alpha 27 + 0$	3.347971
28	0.062742	−0.012742	−0.062737	$\alpha 28 + \pi$	3.128851	$\alpha 28 + 0$	3.128851
29	0.068119	−0.851342	−0.044888	$\alpha 29 + \pi$	2.290251	$\alpha 29 + 0$	2.290251
30	0.066353	−0.991173	−0.036342	$\alpha 30 + \pi$	2.150419	$\alpha 30 + 0$	2.150419

shape is analysed and this may be useful for partially obscured specimens. A small section of our smallest *Asterionella* (Fig. 13) was slightly obscured by the edge of a *Gomphonema* Ehrenberg valve. With good imaging software (such as NIH Image), the whole outline of such valves can be recovered and analysed. However, we could have analysed the valve outline minus the section overlapped by the *Gomphonema* and still obtained shape coefficients.

There are some matters to keep in mind when using this method of Fourier analysis. The Fourier series is used to describe shape in a pointwise fashion, but it may not provide a unique solution. That is, Fourier series terms may or may not converge to zero, but yet have summability (Edwards 1967). In shape analysis a truncated series or a partial sum of terms is used, and these Fourier coefficients completely determine the shape function for a particular outline. Convergence depends on

$$\int_{-\pi}^{\pi} |f(t)|\, dt$$

being finite (Weinberger 1995); then, the shape function, and therefore the Fourier series, converges to $f(t)$ in a pointwise fashion.

Another matter worth considering is choice of a starting point. The shape function is not dependent on any particular starting point. In the simplest closed curve outline, the circle, any point can be chosen to start the tracing of that circle. The same is true for more complex deformations of a circle as represented by the shape function, and therefore Fourier coefficients. These coefficients are invariant to translation, rotation, changes in arc length or dilation (Zahn & Roskies 1972). If a starting point, Z_0, is then chosen for an outline, and a different starting point, Z'_0, is chosen on that same outline, the shift in starting point results in Fourier coefficients as $A'_0 = A_n$ and $\alpha'_n = \alpha_n + n\Delta\alpha$, where $\Delta\alpha = -2\pi\Delta l/L$. That is, the amplitudes are invariant, and there is a shift in phase angle. The same idea can be applied to a group of similar shaped-specimen outlines, such as *Asterionella*, or any other group of similarly shaped diatoms. This would preserve pseudolandmarks for analysis of specimens, particularly those in a size diminution series.

Close inspection reveals that using too many coefficients – in our example, with 26–30 coefficients – produces a distorted outline (Fig. 14, cf. Fig. 1). This overfitting is a result of accumulating error as more coefficients are used (Bennett & MacDonald 1975; Davis 1986). Deciding on the number of coefficients appropriate for analysing and reconstructing a particular outline depends on the number of original coordinates used. In general, the more complex the outline is, the more equally spaced x,y coordinates will be necessary to recover that outline, and the more Fourier coefficients will be necessary to represent the outline. For example, an elliptical shape is recovered by the mean (zero'th) Fourier coefficient, two coefficients produce an ellipse with symmetrical valleys, whereas *Asterionella*'s shape is recovered without waviness in the straight areas by ≥ 16 coefficients (Fig. 14).

It has been suggested that guidance concerning the number of Fourier coefficients that is appropriate in a particular case

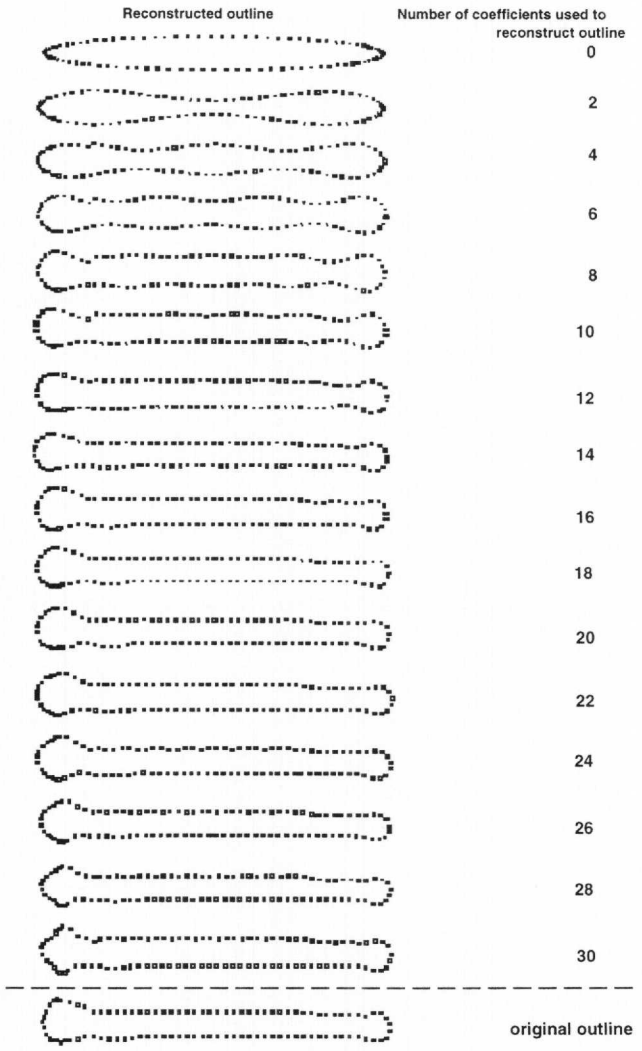

Fig. 14. Series of reconstructed outlines for *Asterionella* specimen 25634b-8 using Fourier coefficients. Outlines using even coefficients are depicted, together with the original outline (at base).

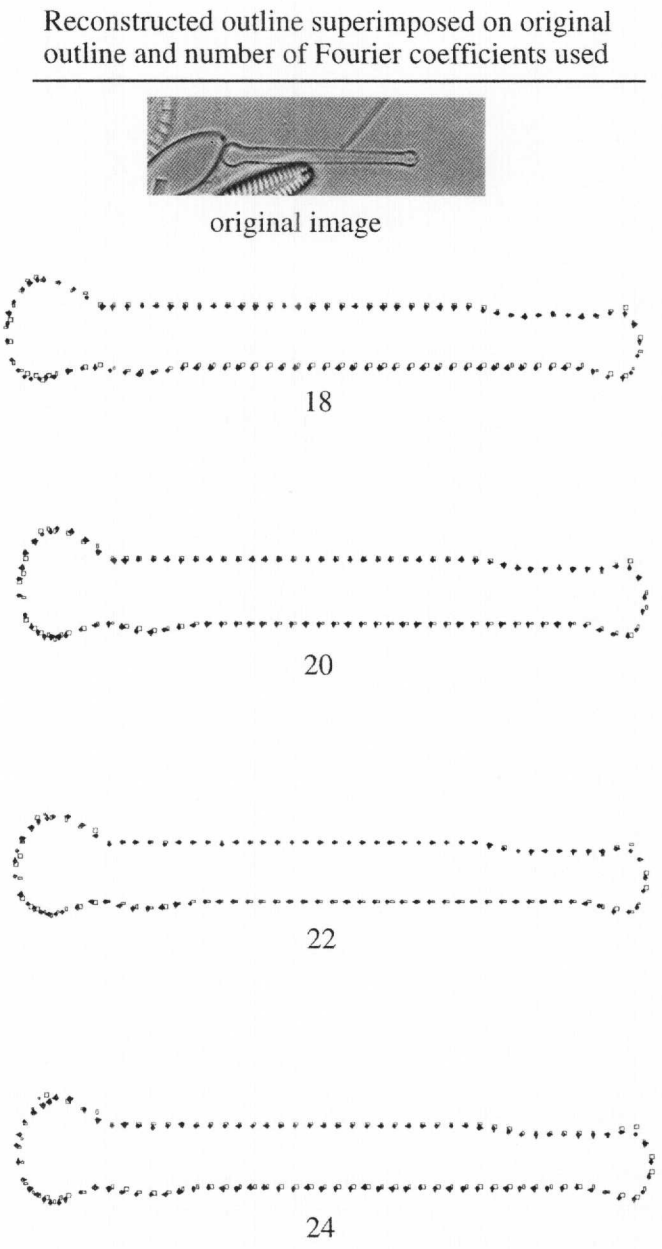

Fig. 15. Comparison of reconstructed outlines (filled symbols) and original outlines (hollow symbols) for *Asterionella* specimen 25634b-8 (length 30.2 μm), based on 100 coordinates and using 18, 20, 22, or 24 Fourier coefficients.

may be gained from spectral analysis and by using the Nyquist frequency (see Introduction: Jenkins & Watts 1968; Koopmans 1974; Platt & Denman 1975; Legendre & Legendre 1983; Davis 1986). For our analysis, 100 *x,y* coordinates were used, based on the Nyquist frequency, and so the number of coefficients to use is 50. However, our empirical tests show that this is not appropriate. No improvement in shape match is evident after 22 coefficients are used, and shape match obviously deteriorates (through overfitting) when 26 or more coefficients are used (Fig. 14). The number of coefficients to use is thus much less than the theoretical limit defined by the Nyquist frequency (Bennett & MacDonald 1975).

During the early stages of a morphometric study, it is necessary, therefore, to carry out empirical tests to determine the number of Fourier coefficients that will be needed for the particular range of shapes encountered. Initially, this can be done by superimposing reconstructed outlines on the original outlines: plots of the data can be resized to view pointwise matching between original *x,y* coordinates and those from the reconstructed outlines. Statistical measures should also be used, as we have described, to determine the relationship between the numbers of coordinates and Fourier coefficients for the diatoms under investigation. This helps to ensure consistency of comparison within a class of shapes with respect to pseudolandmarks. It is best to consider many statistical measures rather than one or a few, as our examples show, and, as more outlines are tested for a given class of closed curves (such as *Asterionella*), it is possible to arrive at a confident estimate of the number of Fourier coefficients necessary to produce a best fit.

The choice of coordinates will also determine how well the reconstructed outline approximates the original shape of the valve. Here, there is a problem of edge detection (Canny 1986). Diatom frustules are composed of hypovalve, epivalve

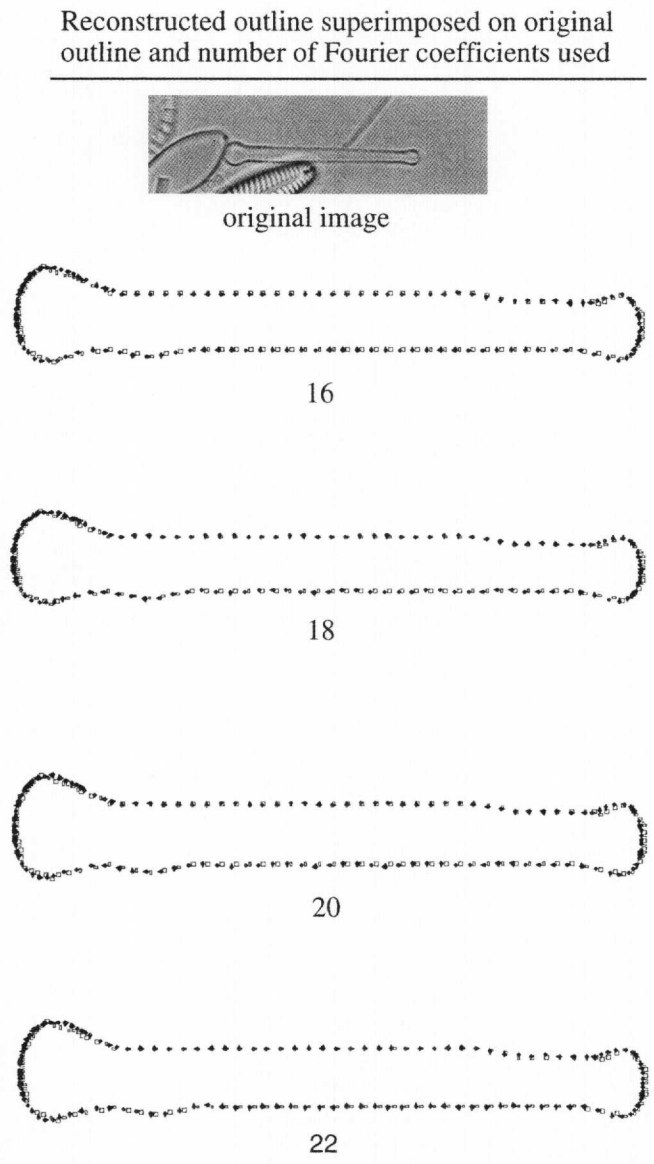

Fig. 16. Comparison of reconstructed outlines (filled symbols) and original outlines (hollow symbols) for *Asterionella* specimen 25634b-8, based on 140 coordinates and using 16, 18, 20, or 22 Fourier coefficients.

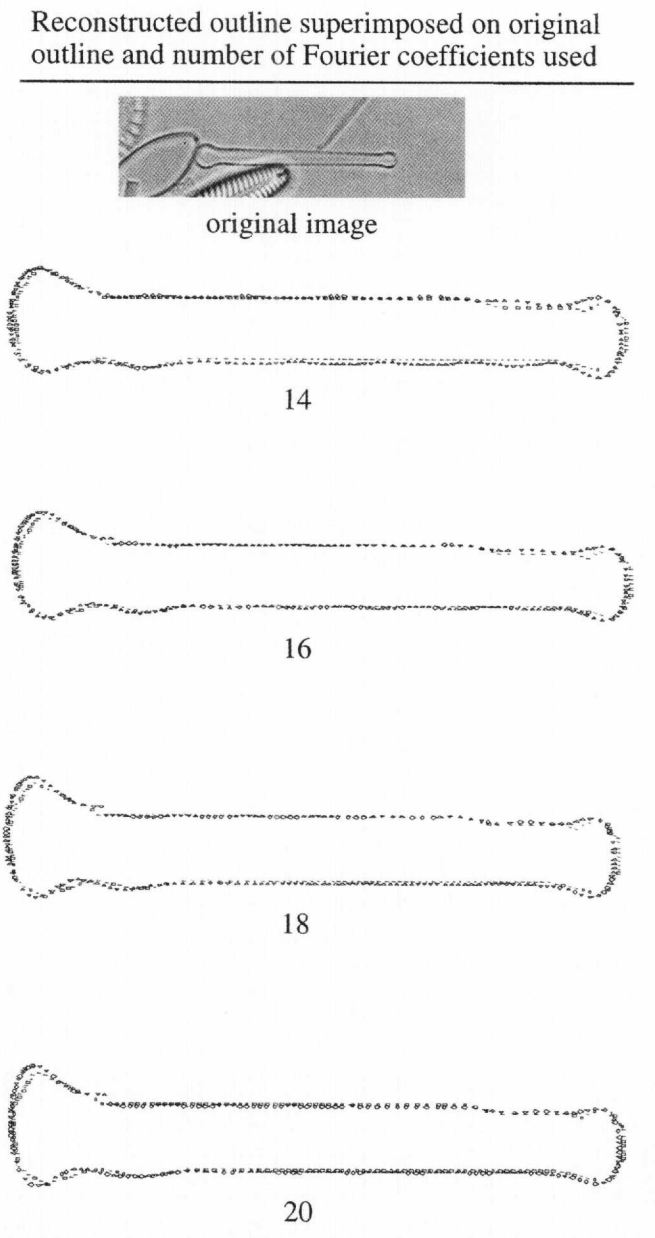

Fig. 17. Comparison of reconstructed outlines (filled symbols) and original outlines (hollow symbols) for *Asterionella* specimen 25634b-8, based on 200 coordinates and using 14, 16, 18, or 20 Fourier coefficients.

and girdle bands (Round *et al.* 1990). The presence or absence of any of these elements will affect the apparent outline of a specimen. And even when single valves are selected, they may not lie entirely flat in a plane. Furthermore, the valve margin is apt not to be imaged as a thin, crisp line, so that determining the true values for *x,y* coordinates is not easily accomplished. It is essential always to check that extracted coordinates do accurately replicate the valve outline, by visual inspection vs the original digitized image.

The results of our statistical analyses indicate that not much is gained in using more coordinates to reduce the number of Fourier coefficients that have to be calculated to reconstruct a best-fit outline for *Asterionella.* The amount of computation time is greater using 140 or 200 coordinates vs 100 coordinates, but the consequent reduction of even Fourier coefficients is only two (from 22 to 20). In studies using a variety of Fourier methods, the number of coordinates used has ranged from 72 for clams (Gevirtz 1976), which have a relatively simple outline, to 256 for maple and sassafras leaves (Kincaid & Schneider 1983), which have very wavy, complex outlines. Likewise, the number of coefficients calculated has ranged from 6 for ostracods (Younker & Ehrlich 1977) and 10 for clams (Gevirtz 1976) to 20 for maple and sassafras leaves (Kincaid & Schneider 1983). Researchers must determine the initial number of coordinates needed to recreate an outline that captures the degree of waviness in that outline.

The number of even Fourier coefficients required to give best fits to *Asterionella* outlines, for a wide range of sizes and using 100 *x,y* coordinates, is 22 (Tables 4, 7, 8: for the 51.9 μm specimen, 20 coefficients might also be sufficient) and

Table 4. Pointwise comparison between x coordinates, y coordinates, and distance between x,y coordinates for original outline and reconstructed outlines based on 18, 20, 22, and 24 coefficients for *Asterionella* specimen 25634b-8, 30.21 mm length. One hundred coordinates were used.

	x	y	Distance
18 Coefficients			
Average	−0.04053	−0.21433	0.15925
Variance	0.01071	0.63184	0.00506
Standard deviation	0.10349	0.79488	0.07114
Coefficient of variance	−2.55365	−3.70869	0.44670
20 Coefficients			
Average	−0.02197	−0.22381	0.14351
Variance	0.00978	0.62729	0.00548
Standard deviation	0.09888	0.79201	0.07404
Coefficient of variance	−4.50100	−3.53871	0.51594
22 Coefficients			
Average	−0.00937	−0.22667	0.14080
Variance	0.00908	0.61751	0.00558
Standard deviation	0.09529	0.78582	0.07468
Coefficient of variance	−10.17462	−3.46680	0.53037
24 Coefficients			
Average	−0.01036	−0.21756	0.16315
Variance	0.00902	0.59435	0.00513
Standard deviation	0.09498	0.77094	0.07164
Coefficient of variance	−9.16451	−3.54352	0.43909

this is the best choice to cover the size range for *Asterionella*. Application of the statistical measures we used does not provide unequivocal guidance on how many coefficients will be appropriate but augments empirical use of plots of reconstructed outlines to decide best fit. It is important to avoid overfitting, which can easily occur in smaller specimens compared to larger specimens in a size diminution series, and sta-

Table 5. Pointwise comparison between x coordinates, y coordinates, and distance between x,y coordinates for original outline and reconstructed outlines based on 16, 18, 20, and 22 coefficients for *Asterionella* specimen 25634b-8, 30.21 mm length. One hundred forty coordinates were used.

	x	y	Distance
16 Coefficients			
Average	−0.09761	−0.00009	0.17734
Variance	0.00800	0.03567	0.01256
Standard deviation	0.08944	0.18887	0.11205
Coefficient of variance	−0.91632	−2145.03109	0.63183
18 Coefficients			
Average	−0.30790	0.03238	0.16649
Variance	0.27526	0.00390	0.01338
Standard deviation	0.52465	0.06246	0.11565
Coefficient of variance	−1.70395	1.92876	0.69466
20 Coefficients			
Average	−0.09725	0.02929	0.17373
Variance	0.00745	0.00388	0.01450
Standard deviation	0.08632	0.06228	0.12043
Coefficient of variance	−0.88761	2.12605	0.69318
22 Coefficients			
Average	−0.03584	0.02674	0.17319
Variance	0.00182	0.00317	0.01492
Standard deviation	0.04270	0.05634	0.12215
Coefficient of variance	−1.19152	2.10711	0.70526

Table 6. Pointwise comparison between x coordinates, y coordinates, and distance between x,y coordinates for original outline and reconstructed outlines based on 14, 16, 18, and 20 coefficients for *Asterionella* specimen 25634b-8, 30.21 mm length. Two hundred coordinates were used.

	x	y	Distance
14 Coefficients			
Average	−0.06522	0.09747	0.25081
Variance	0.00954	0.01496	0.07290
Standard deviation	0.09767	0.12232	0.26999
Coefficient of variance	−1.49771	1.25492	1.07648
16 Coefficients			
Average	−0.03048	0.06643	0.23675
Variance	0.00515	0.01304	0.07124
Standard deviation	0.07175	0.11420	0.26691
Coefficient of variance	−2.35406	1.71903	1.12735
18 Coefficients			
Average	−0.01229	0.03889	0.71863
Variance	0.00495	0.01640	0.07073
Standard deviation	0.07035	0.12805	0.26595
Coefficient of variance	−5.72576	3.29292	0.37008
20 Coefficients			
Average	−0.00402	0.02770	0.26068
Variance	0.00523	0.01813	0.06882
Standard deviation	0.07233	0.13466	0.26233
Coefficient of variance	−18.00826	4.86080	1.00631

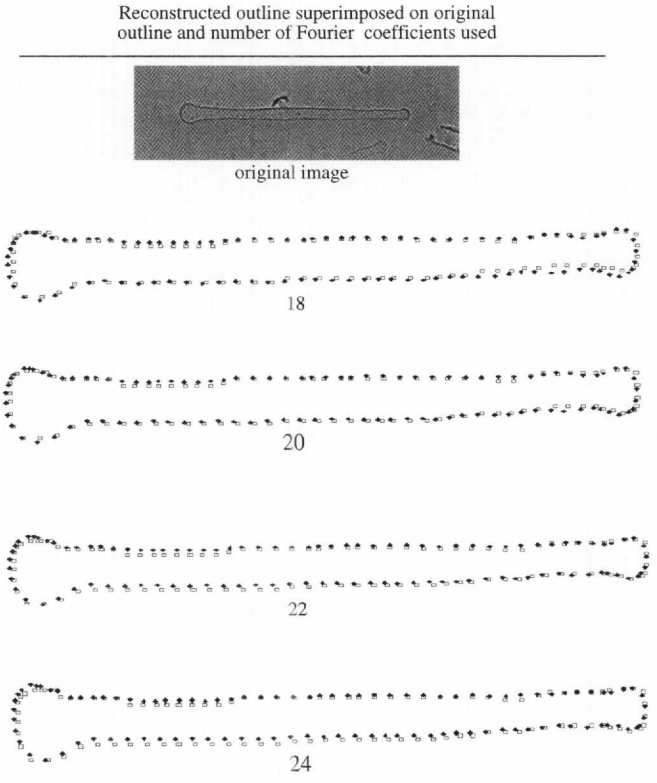

Fig. 18. *Asterionella* specimen 25634b-10 (51.9 μm): comparison of reconstructed outlines (filled symbols) and original outlines (hollow symbols), based on 100 coordinates and using 18, 20, 22, or 24 Fourier coefficients.

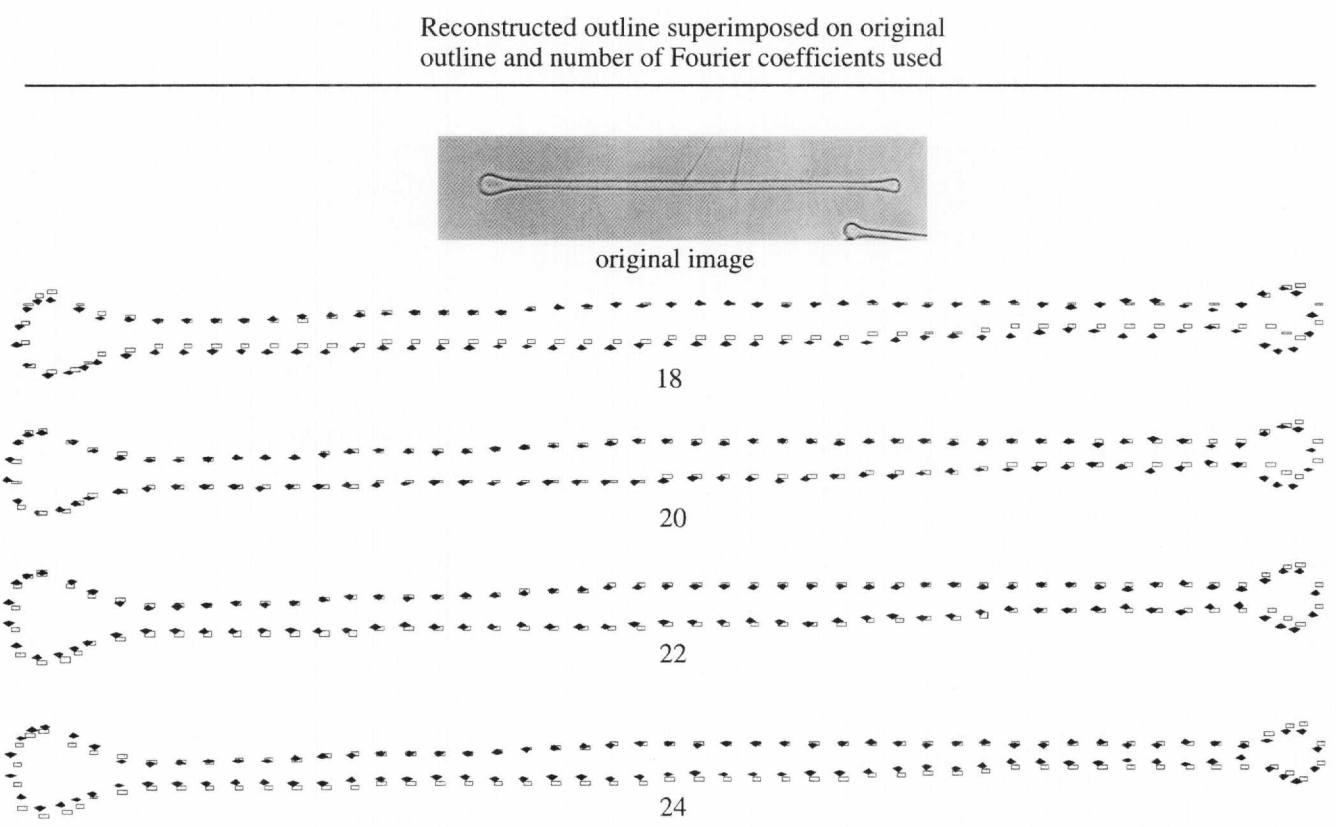

Fig. 19. *Asterionella* specimen 1997-1 (94.6 μm): comparison of reconstructed outlines (filled symbols) and original outlines (hollow symbols), based on 100 coordinates and using 18, 20, 22, or 24 Fourier coefficients.

Table 7. Pointwise comparison between *x* coordinates, *y* coordinates, and distance between *x,y* coordinates for original outline and reconstructed outlines based on 18, 20, 22, and 24 coefficients for *Asterionella* specimen 25634b-10, 51.88 mm length. One hundred coordinates were used.

	x	*y*	Distance
18 Coefficients			
Average	0.02259	−0.02440	0.07172
Variance	0.00033	0.00225	0.00154
Standard deviation	0.01812	0.04739	0.03925
Coefficient of variance	0.80205	−1.94272	0.54733
20 Coefficients			
Average	0.04823	0.00646	0.06710
Variance	0.00131	0.00129	0.00083
Standard deviation	0.03623	0.03596	0.02880
Coefficient of variance	0.75117	5.56705	0.42921
22 Coefficients			
Average	0.06891	0.02539	0.06768
Variance	0.00275	0.00109	0.00081
Standard deviation	0.05242	0.03301	0.02840
Coefficient of variance	0.76081	1.30040	0.41962
24 Coefficients			
Average	0.08421	0.03333	0.07018
Variance	0.00425	0.00116	0.00101
Standard deviation	0.06521	0.03407	0.03184
Coefficient of variance	0.77433	1.02238	0.45371

Table 8. Pointwise comparison between *x* coordinates, *y* coordinates, and distance between *x,y* coordinates for original outline and reconstructed outlines based on 18, 20, 22, and 24 coefficients for *Asterionella* specimen 1997-1, 94.58 mm length. One hundred coordinates were used.

	x	*y*	Distance
18 Coefficients			
Average	−0.01036	−0.21756	0.16315
Variance	0.00902	0.59435	0.00513
Standard deviation	0.09498	0.77094	0.07164
Coefficient of variance	−9.16451	−3.54352	0.43909
20 Coefficients			
Average	−0.09780	−0.05293	0.10866
Variance	0.00649	0.00471	0.00607
Standard deviation	0.08057	0.06862	0.07793
Coefficient of variance	−0.82382	−1.29656	0.71719
22 Coefficients			
Average	−0.05378	−0.00151	0.09782
Variance	0.00224	0.00590	0.00364
Standard deviation	0.04730	0.07680	0.06034
Coefficient of variance	−0.87943	−50.76958	0.61689
24 Coefficients			
Average	−0.02310	0.03989	0.11115
Variance	0.00123	0.01047	0.00403
Standard deviation	0.03508	0.10231	0.06348
Coefficient of variance	−1.51851	2.56485	0.57114

tistical analysis is helpful in determining the number of *x,y* coordinates to use within a size diminution series to preserve pseudolandmarks. In other pennate diatoms, valves vary greatly in shape; at one extreme are simple elliptical forms, such as in *Diploneis* Ehrenberg, whereas at the other extreme are some with a high degree of waviness, such as *Eunotia* Ehrenberg, and these will require different combinations of coordinate and Fourier coefficients for morphometric analysis. Once one has decided on the shape class of the specimens that are to be included in an analysis, the following questions must be answered in sequence. First, what is the size range of specimens for analysis? Second, how many *x,y* coordinates are necessary to recreate the outline of the smallest specimen when superimposed on its digitized outline? Third, can the outline of the largest specimen be recreated with the same number of coordinates used for the smallest specimen (to preserve pseudolandmarks)? Then, once the number of *x,y* coordinates to use has been determined, the number of Fourier coefficients necessary to reconstruct a best-fit outline will need to be determined. So, for the smallest specimen, calculate enough Fourier coefficients for each of the even harmonics until an approximate outline is achieved and overfitting is not evident (by reconstructing outlines). Use the statistical measures to determine which of the reconstructed outlines (with respect to the number of Fourier coefficients used) is closest to the original. Repeat for selected valves over a range of sizes and determine a consensus number of coefficients that can be applied throughout.

Fourier coefficients, as well as coefficients from other types of orthogonal polynomial regression, can be used to characterize shape similarities for taxonomic purposes. Multivariate statistical methods, including principal components analysis (e.g. Stoermer & Ladewski 1982; Lohmann 1983), discriminant analysis (Younker & Ehrlich 1977), and cluster analysis (Ferson *et al.* 1985), as well as frequency distributions (Kincaid & Schneider 1983) have been used to analyse shape descriptors. Multivariate statistical methods are especially applicable in analyses of Fourier coefficients, since these coefficients are orthogonal (see Methods). As each pair of Fourier coefficients, amplitude and phase angle, is calculated, they are inserted into the Fourier expansion of the shape function,

$$\Phi^*(t) = a_0 + \sum_{n}^{N} A_n \cos(nt + \alpha_n),$$

and summed with previous results. One summed quantity can be used to reconstruct an outline independent of subsequent summed quantities in the Fourier expansion. Orthogonality means independence between pairs of Fourier coefficients, and ordination from multivariate statistical methods depicts this independence as different attributes of diatom valve shape being assigned to different dimensions (eigenvectors) in shape space. Moreover, orthogonality means that two similar shapes will have similar shape descriptors and will be depicted near each other in ordination diagrams of shape space. Quantification of shape descriptors is therefore useful in taxonomic sorting and classification (e.g. Younker & Ehrlich 1977; Stoermer & Ladewski 1982; Stoermer *et al.* 1984).

ACKNOWLEDGEMENTS

We thank Ted Ladewski for his pioneering work using Legendre polynomials in diatom taxonomy. Because of his work, impetus for the application of other orthogonal polynomials, such as Fourier analysis, was provided. In addition, we give special thanks to Dr Regine Jahn, Botanischer Garten und Botanisches Museum Berlin-Dahlem, Germany, for access to the H. Körner collection. This research was supported by NSF Grant DEB(PEET) 9521882.

REFERENCES

BENNETT J.R. & MACDONALD J.S. 1975. On the measurement of curvature in a quantitized environment. *IEEE Transactions on Computers* 24: 803–820.

BROWN J.W. & CHURCHILL R.V. 1993. *Fourier series and boundary value problems.* McGraw-Hill, New York. 348 pp.

CANNY J. 1986. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8: 679–698.

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

EDWARDS R.E. 1967. *Fourier series: a modern introduction,* vol. 1. Holt, Rinehart & Winston, New York. 208 pp.

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

GEVIRTZ J.L. 1976. Fourier analysis of bivalve outlines: implications on evolution and autecology. *Mathematical Geology* 8: 151–163.

GOLDMAN N., PADDOCK T.B.B. & SHAW K.M. 1990. Quantitative analysis of shape variation in populations of *Surirella fastuosa. Diatom Research* 5: 25–42.

JACKSON D. 1941. *Fourier series and orthogonal polynomials.* Carus Mathematical Monographs. No. 6. Mathematical Association of America. Washington, DC. 234 pp.

JENKINS G.M. & WATTS D.G. 1968. *Spectral analysis and its applications.* Holden-Day, San Francisco. 525 pp.

KINCAID D.T. & SCHNEIDER R.B. 1983. Quantification of leaf shape with a microcomputer and Fourier transform. *Canadian Journal of Botany* 61: 2333–2342.

KOOPMANS L.H. 1974. *The spectral analysis of time series.* Academic Press, New York. 366 pp.

KUFNER A. & KADLEC J. 1971. *Fourier series.* Iliffe Books, London. 358 pp.

KUHL F.P. & GIARDINA C.R. 1982. Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* 9: 3236–3258.

LEGENDRE L. & LEGENDRE P. 1983. *Numerical ecology: developments in modelling,* vol. 3. Elsevier Scientific Publishing, Amsterdam. 419 pp.

LOHMANN G.P. 1983. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. *Mathematical Geology* 15: 659–672.

MOU D. & STOERMER E.F. 1992. Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors. *Journal of Phycology* 28: 386–395.

PERSOON E. & FU K.S. 1977. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics* 7: 170–179.

PLATT T. & DENMAN K.L. 1975. Spectral analysis in ecology. *Annual Review of Ecology and Systematics* 6: 189–210.

ROHLF F.J. 1990. Fitting curves to outlines. In: *Proceedings of the Michigan Morphometrics Workshop, Special Publication 2* (Ed. by F.J. Rohlf & F.L. Bookstein), pp. 167–177. The University of Michigan Museum of Zoology, Ann Arbor, Michigan.

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

ROUND F.E., CRAWFORD R.M. & MANN D.G. 1990. *The diatoms: biology & morphology of the genera.* Cambridge University Press, Cambridge. 747 pp.

STEINMAN A.D. & LADEWSKI T.B. 1987. Quantitative shape analysis

of *Eunotia pectinalis* (Bacillariophyceae) and its application to seasonal distribution patterns. *Phycologia* 26: 467–477.

STOERMER E.F. 1996. A simple, but useful, application of image analysis. *Journal of Paleolimology* 15: 111–113.

STOERMER E.F. & LADEWSKI T.B. 1982. Quantitative analysis of shape variation in type and modern populations of *Gomphoneis herculeana*. *Nova Hedwigia* 73: 347–386.

STOERMER E.F., LADEWSKI T.B. & KOCIOLEK J.P. 1984. Further observations on *Gomphoneis*. In: *Proceedings of the Eighth International Diatom Symposium* (Ed. by M. Ricard), pp. 205–213. Koeltz Scientific Books, Koenigstein.

STOERMER E.F., QI Y.-Z. & LADEWSKI T.B. 1986. A quantitative investigation of shape variation in *Didymosphenia* (Lyngbye) M. Schmidt (Bacillariophyta). *Phycologia* 25: 494–502.

SZOKEFALVI-NAGY B. 1965. *Introduction to real functions and orthogonal expansions*. Oxford University Press, New York. 447 pp.

THERIOT E. & LADEWSKI T.B. 1986. Morphometric analysis of shape of specimens from the neotype of *Tabellaria flocculosa* (Bacillariophyceae). *American Journal of Botany* 73: 224–229.

WEINBERGER H.F. 1995. *A first course in partial differential equations*. Dover Publications, New York. 446 pp.

YOUNKER J.L. & EHRLICH R. 1977. Fourier biometrics: harmonic amplitudes as multivariate shape descriptors. *Systematic Zoology* 26: 366–342.

ZAHN C.T. & ROSKIES R.Z. 1972. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers* 21: 269–281.

Accepted 22 June 2001.