



Quantitative morphometric methods in diatom research

Janice L. Pappas¹, J. Patrick Kociolek² and Eugene F. Stoermer^{3†}

¹ Museum of Paleontology, University of Michigan, 1109 Geddes Ave., Ann Arbor, MI 48109-1079, USA; jlpappas@umich.edu

² Department of Ecology and Evolutionary Biology, University of Colorado at Boulder, Boulder, CO 80309-0218, USA; patrick.kociolek@colorado.edu

³ School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109-1041, USA; †deceased

Abstract: Morphometric methods have been used in diatom research for decades. We present a review of the history of usage of morphometric methods of outline shape analysis, pattern recognition, and landmark-based analysis. In addition, we present how morphometric methods are important in diatom taxonomy and classification and what connections exist between morphometric methods and biologically meaningful results. Next, we present some details about calculating shape descriptors and using them in analysis of shape variation, the issues to be aware of, and what such results mean when defining shape groups as species groups. Finally, we provide a glimpse of the future in using morphometric methods in diatom research.

Key words: Diatoms, morphometrics, shape analysis, Legendre coefficients, Fourier transform, landmarks, outline analysis

Introduction

With a shift from purely descriptive works to measurement-based studies, morphometric methods have been used to help separate diatom taxonomic groups which are difficult to differentiate by visual means alone. Moreover, because of the laborious task of counting diatoms for aquatic and ecological studies as well as paleontological investigations, automation of diatom identification has been a desirable extension of quantitative taxonomic methods. Development of such methods has progressed from simple measurements to more mathematically involved treatments, such as shape analysis and pattern recognition.

Variation in diatoms has been the focus of many studies over the long history in the study of these organisms. With the advent of the microscope, and subsequently, the acceptance of Darwinian ideas on evolution and change, diatom studies progressed as research encompassed an increasing interest in the details of morphological features. At a time when diatom research noticeably advanced, there were a large number of species descriptions and studies analyzing the range in a variety of metrics associated with diatom cell walls (e.g., length, breadth, striae density, and the absolute number and/or density of a wide array of valve features; see e.g., Reimer 1954).

Over 35 years ago, more of a biometric or traditional morphometric approach to diatom variation was carried out. Burke (1970) plotted the relationship between valve diameter and size in *Aulacodiscus* species, noting the variation associated with diameter, and the variation that tended to be independent of size. Other investigators, trying to understand the potential impact of environmental variables on valve size and structure pursued analyses with cultured cells, noting the relationship of valve size and ornamentation of diatom cell walls with a range of environmental variables (e.g., Geissler 1970a, b, 1982). A large number of studies then followed in this tradition, plotting a number of diatom metrics to establish variation and possible implications of the variation (e.g., Holmes & Reimann 1966, Steinman & Sheath 1984, Kociolek & Stoermer 1988, Rivera & Barrales 1994, Kobayashi et al. 1998, Wendker 1990). Others applied statistics to assess the probability of the sources of variation (Wood et al. 1987,

Mizuno 1987). Multivariate statistical analyses were applied to structural variation in, mainly, centric diatoms (Genkal & Kuzmin 1979, Theriot & Stoermer 1982, 1984, Theriot 1987, Theriot et al. 1988), to assess variation and document cladogenesis in diatom lineages (Theriot 1992). Tropper (1975) had applied statistical analysis to document differences in natural populations of *Achnanthes hauckiana* Grunow in Cleve & Grunow and to formulate shape differences based on size. Investigations also focused on the types of data that might be useful in these types of analyses (e.g., Johansen & Theriot 1987).

All of the methods used have been successful to varying degrees in quantifying differences among taxa. When speaking of morphometrics as a discipline, quantitative morphometrics is the focus of subject-matter that includes outline (Slice et al. 1996) and landmark-based (Slice et al. 1996) methods, although there is a trend currently toward combining the tools of outline and landmark-based (geometric) morphometrics into new methods. Degree of success in using all these methods depends, in part, on the actual applicability of the method chosen for analysis. Moreover, results from morphometric analysis are the most convincing when they encompass biological meaning rather than being an exercise in mathematical application only. Biological meaning is accomplished by results based on landmark-based methods that are choice of points on the interior of organisms and represent homologous points on homologous structures, or outline methods which depend on choice of points on the exterior of organisms and represent biological change in shape.

Purposes of this review

Traditional morphometrics (Slice et al. 1996) continue to be useful in diatom research (e.g., Stoermer & Andresen 2006), but we will not devote further attention to this topic at this time. We are interested in, and focus on, the more sophisticated mathematical methods in morphometrics as useful tools in diatom research with respect to the advanced mathematical treatments of morphometrics, and where relevant, pattern recognition studies that rely on the mathematics of morphometrics. A compilation of the advanced mathematical methods used, the validity of their use in research studies, and their impact on diatom taxonomy may be assessed.

To this end, our review will be composed of four parts. After the introductory remarks already made, the first section is devoted to the history of the theoretical concepts and bases behind quantitative morphometrics that are relevant to the development of methods for use in diatom research. Without a theoretical basis to encompass the outcome of particular methods, the context of and potential for understanding such methods in diatom research is absent. This context involves the concept of shape space, two general bases of analyzing shape, namely, landmark-based and outline-based methods, and multivariate methods used to analyze the shape descriptors. The primary, relevant literature will be cited so that researchers may consult each topic in more depth.

In the second section, a history of using morphometric methods in diatom research is presented in chronological order. Since this is a review about methods and how those methods are applied, an assessment is presented of using morphometrics in this subsection. We provide an assessment of landmark-based methods in this section, and treat an assessment of other morphometric methods, namely, shape analysis, in the next section. Because of the diatom life cycle, especially size reduction during vegetative reproduction, characters may be lost and the application of geometric morphometrics may present some challenges. In addition, we elucidate some basic concepts that are necessary to truly master and effectively apply landmark-based methods in diatom research.

In the third section, outline shape analysis methods are specifically covered since these methods have been used the most. In addition, we present an assessment of these methods in their capacity to recover biologically meaningful features of diatoms. We cover seven subtopics: Diatom valve shape as a morphological character; Basic measurements; Outline shape, orthogonal polynomials and their properties, and regression coefficients; Shape descriptors and data

transformation; Multivariate methods and the analysis of shape variation (including principal components and discriminant analysis); Choice of method and issues to consider; Shape descriptors, shape space and biologically meaningful interpretation.

In the final section, potential application of combination and new geometric morphometric methods are considered. Much has been accomplished and advanced in quantitative morphometrics recently, and applicability in diatom research may very well profit from these advances.

With these purposes in mind, we will detail the basis, chronology and usage of quantitative morphometric methods in diatom research. At this point, some generalities are in order. To better understand the terms of geometric morphometrics, see Slice et al. (1996) for a glossary. For a general primer on geometric morphometrics, see [Zelditch et al. \(2004\)](#). For background on shape analysis, see [Stoermer & Ladewski \(1982\)](#) and [Pappas & Stoermer \(2003\)](#) for information concerning Legendre polynomials, and see [Mou & Stoermer \(1992\)](#) and [Pappas et al. \(2001\)](#) for information concerning Fourier analysis in diatom research.

Theoretical context of quantitative morphometric methods in diatom research

During the late 1970s, development of morphometric methods as applied to biological problem-solving gained interest and prominence (e.g., [Bookstein 1978](#), [Kendall 1977](#)). In particular, shape analysis was seen to provide a quantitative way to characterize biological organisms that were empirically difficult to differentiate. This was seen as especially promising for taxonomic resolution at the species level.

Two different but complementary approaches were devised on how to represent shapes. One approach involved using points common to all specimens whereby those points formed a continuum (e.g., [Bookstein 1991](#)). These points could then be compared to a mean shape (e.g., [Goodall 1991](#)) for the group of specimens being considered. The points are used for all specimens insofar as the change in shape is measured by the deformation among all shapes. The other approach involved using points common to all specimens whereby the lowest dimensional space was achieved ([Kendall 1977, 1984, Kendall et al. 1999](#)). From this, only those specimens where all points do not align completely are considered. The points determine the object of study ([Kendall et al. 1999](#)). The choice of points as markers or codes of the shape of a specimen in relation to all other specimens being considered is not a trivial task ([Kendall et al. 1999](#)). Others involved with morphometrics, both early on (e.g., [Ziezold 1977, 1994](#)), and subsequently (e.g., [Mardia & Dryden 1989a, 1989b](#), [Dryden & Mardia 1991, 1998](#), [Small 1988, 1996](#)) used aspects of both approaches.

Inherent in [Kendall et al. \(1999\)](#) approach is the idea of a shape space (Slice et al. 1996), where points from all specimens may be rotated or translated and are dispersed with respect to their centroid being moved to the origin of the coordinates. Size is expanded or contracted to yield a standardized plot, and this plot is a result in a configuration known as pre-space or preshape space. [Kendall \(1984\)](#) and [Kendall et al. \(1999\)](#) were concerned with shape manifolds and complex projective spaces (also see e.g., [Mardia & Patrangenaru 2005](#)). While this approach is theoretically useful as a basis for morphometrics, biological shapes are more suited to being studied using continuous deformation methods ([Bookstein 1991](#)).

The statistical variation of size and shape in organisms is front and center of [Bookstein's \(1986, 1991\)](#) approach. To reiterate, distribution of shape variation around mean shapes is the basis of the approach. That is, the tangent space in the neighborhood of a particular point in shape space is the tangent projection that is a linearized version of the average shape. The choice of points, also known as landmarks (Slice et al. 1996), are important in allometric variation among shapes, both globally and locally ([Dryden & Mardia 1998](#)). Landmarks are categorized depending on how the points are chosen. Points chosen on an organism that correspond to those on another organism in a biologically important way are anatomical landmarks. Points chosen with reference to maximum curvature are mathematical landmarks. Points chosen as maximum

diameters, centroids, foci, or other extrema are pseudolandmarks (Slice et al. 1996). Points chosen on a curve with respect to change in angle along the curve in low dimensional space are semilandmarks (Dryden & Mardia 1998).

Landmarks may be analyzed by statistically transforming them into shape coefficients. Translation, rotation and rescaling must be removed (Rohlf & Slice 1990) to achieve matching of shapes. This superimposition (Slice et al. 1996) is used in order to estimate an average shape. Once an average shape is determined, configurations of landmarks among shapes may be analyzed using Procrustean distances (Gower 1975). Matching of shapes is obtained by least-squares regression. Whether using shape coefficients or Procrustean distances, the results may be analyzed by multivariate statistics (e.g., Mardia & Dryden 1989a, 1989b) and displayed in shape space (e.g., Goodall & Mardia 1993). In Kendall's treatment (1984, Kendall et al. 1999), shape space is a quotient space whereby a geodesic distance is a Riemannian metric with a Fréchet mean space (Le & Kume 2000). Comparing two shapes in this way may be approximately a Procrustean mean from minimization of Euclidean distances (Kendall 1984, Kendall et al. 1999) when small variation among shapes is present (Slice 2001).

Expressing allometric variation with respect to the comparison of forms as a latent variable originates from the work of Thompson (1917, 1942), and this basis is used in modern morphometrics (e.g., Adams et al. 2004). The multivariate method most often chosen to depict shape variability is principal components analysis (PCA). Although size and shape are converted to orthogonal variables, size cannot be entirely extricated from the analysis (Mosimann 1970). An alternative use of PCA in the analysis of shape variation involves warping based on thin-plate splines (Bookstein 1989). Mapping one configuration of points to another involves two parts to the decomposition of the deformation (= bending) of shapes. Tilting an infinitely thin plate so that bending energy (Slice et al. 1996) equals zero produces an affine transformation on a global scale. The latent roots or eigenvalues of the bending energy matrix corresponding to the bending energy in the non-affine part of the transformation are on a local scale. The latent vectors or eigenvectors associated with the latent roots are the important part of the decomposition since they are biologically interpretable with respect to the organisms being studied. The eigenvalues are bending energies, and the eigenvectors are principal warps (Dryden & Mardia 1998).

History of morphometric methods in diatom research – a chronology

Geometric morphometrics, and outline shape analysis in particular, was viewed as a way to quantitatively differentiate morphologically similar specimens that form species complexes or other taxonomic units or as a means to pattern recognition for purposes of counting diatoms in limnological and ecological studies. In the area of pattern recognition, the desire for automated taxon identification systems has been an impetus for the improvement of consistency and efficacy in counting diatoms and is directly linked to and a result of using such morphometric methods (MacLeod et al. 2007).

One of the earliest attempts at morphometric analysis in diatom research involved automated identification of diatoms using holographic filters (Cairns, Jr. et al. 1974, 1977, 1982). The idea involved using a laser light source and superimposition of a diatom image and a plane wave that were focused on a photographic plate. The filter recorded the optical frequencies of the diatom being studied (Almeida & Eu 1976). In this pattern recognition system, when a diatom corresponded to the holographic filter made of that diatom, the intensity of the signal was as indication of degree of identification. The more intense the signal, the higher the correlation was with the correct identification. That is, autocorrelation signified correct identification, while cross-correlation signified misidentification. Composite holographic filters were used to identify multiple diatoms at the same time and diatoms of the same species that had different sizes. Partial diatom frustules could also be identified using holographic filters. A micro-optical processor was used to automate this process for diatom counting. Error rate for identifications was 5–10 %

when compared to counts made visually by a diatom taxonomist using a microscope (Cairns et al. 1982).

At approximately the same time, a preliminary study was conducted on an automated method of identifying phytoplankton (Johnston & Stoermer 1976). Diatoms were used to illustrate the procedure that involved photographing and scanning specimens. Using absorbance values from the scanning process, a topological method was incorporated into computer programs to detect individual cell boundaries. The results were that diatom cells could be distinguished from their background, circular diatoms could be distinguished from oblong diatoms, the ratio of length to width for each cell could be estimated, an asymmetry factor could be calculated, and measurement of cell area could be accomplished. Fascicle frequency, based on Fourier analysis to calculate rotational symmetry, was determined as a way to institute pattern recognition of different centric diatoms. With the problems associated with cell overlap and boundary detection, Johnston & Stoermer (1976) recognized that automated diatom taxonomy is much more difficult to achieve than developing a computer-driven database as a taxonomic reference standard.

In the winter of 1980, The University of Michigan Museum of Zoology hosted a morphometrics workshop and produced two volumes of methods and applications (Bookstein et al. 1985, Rohlf & Bookstein 1990). In general, a consensus was reached that espoused the view that wherever possible, landmarks, as biological homologies, should be used in morphometrics to confer biological meaning to the results of analysis. The group did not address problems where landmarks are not identifiable on organism form because of developmental or other morphological considerations. Outline recovery methods as decomposition into harmonics (e.g., Rohlf 1990) rather than superimposition were also explicated. Diatoms, per se, were one of many organisms not tested with morphometric methods developed at the workshop

Just prior to development of modern shape theory and usage of landmark-based methods, decomposition into harmonics, or Fourier analysis, was developed as a shape analysis method of object outlines (Zahn & Roskies 1972, Bennett & MacDonald 1975, Persoon & Fu 1977). Concomitantly and subsequently, Fourier analysis has become one of the most widely used techniques in calculating biological shape descriptors for use in multivariate ordination or other analyses (e.g., Gevirtz 1976, Younker & Ehrlich 1977, Kincaid & Schneider 1983, Rohlf & Archie 1984, Ferson et al. 1985, Lestrel 1997, and others), including eigenshape analysis (Lohmann 1983) and Fourier elliptical analysis (Kuhl & Giardina 1982). Applications in diatom research would take the form of not only using Fourier analysis, and in particular, the discrete Fourier transform, but also adaptation of orthogonal polynomials for use in shape analysis.

Such methods were being explored with respect to diatom valve shape outlines. One of the first of these studies was that of *Gomphoneis herculeana* (Ehrenberg) P. T. Cleve (Stoermer & Ladewski 1982). Shape analysis of this taxon was accomplished using Legendre polynomials fitted to the diatom outline in a least-squares sense. Legendre coefficients were extracted as shape descriptors and ordinated in PCA to depict shape variation. As a result, *G. herculeana* was determined to have different shapes between a modern and the type population, and this was corroborated using discriminant analysis (DA).

Soon thereafter, other studies using Legendre polynomials followed, including those of *Gomphoneis* P. T. Cleve (Stoermer et al. 1984), *Didymosphenia* M. Schmidt (Stoermer et al. 1986), *Tabellaria flocculosa* (Roth) Kützing (Theriot & Ladewski 1986), *Eunotia pectinalis* (Kützing) Rabenhorst (Steinman & Ladewski 1987), *Surirella fastuosa* (Ehrenberg) Ehrenberg (Goldman et al. 1990), *Meridion* C. A. Agardh (Rhode et al. 2001), *Cymbella cistula* (Hemprich in Hemprich & Ehrenberg) Kirchner species complex (Pappas & Stoermer 2003), and *Fragilariforma* (Ralfs) Williams et Round (Kingston & Pappas 2009). In each of these studies, Legendre coefficients were used as shape coefficients, and these shape coefficients were ordinated in PCA. DA was also used, and with this, ordination revealed that eigenvectors could be interpreted in biological terms. Fourier analysis was also applied in two cases. One involved a study of *Tabellaria* Ehrenberg (Mou & Stoermer 1992), and the other study involved *Asterionella* Hassall (Pappas & Stoermer 2001, Pappas et al. 2001). Both *Tabellaria* and *Asterionella* are particularly devoid

of discrete morphological characters and are prime examples where landmark-based methods are not applicable.

For studies using Legendre polynomials, results indicated varying degrees of distinguishability of shape groups. A broader study of *Gomphoneis* revealed that five of six modern populations were different in shape from the type population. The one population that was similarly shaped occurred in the same locality as the type population (Stoermer et al. 1984). For *Didymosphenia*, populations from Lake Baikal and China could be distinguished from Karluk Lake and Lake Superior when based on shape analysis alone, while populations from Karluk Lake and Lake Superior were not easily distinguishable from each other (Stoermer et al. 1986). Shape variation in neotype specimens of *Tabellaria flocculosa* was found to be separable for long specimens, while size-reduced specimens exhibited greater shape variability, and therefore, shape was not recovered as well (Theriot & Ladewski 1986). In studying populations from a pool site and a riffle site in Rhode Island over a four-season period, *Eunotia pectinalis*, was found to have continuous shape variation (Steinman & Ladewski 1987). Similarly, populations of *Surirella fastuosa* from Belize and the Philippines revealed only slight shape differences (Goldman et al. 1990). Separation of shapes in the *Cymbella cistula* species complex from specimens in northern Lake Michigan and Torch Lake was accomplished to show that six groups could be distinguished (Pappas & Stoermer 2003). In a study of *Fragilariforma* that included specimens from a variety of localities as well as digitized images of specimens published in the literature, it was demonstrated that nine different taxon groups were distinguished and that currently used taxonomic names could be assigned to each group (Kingston & Pappas 2009).

Again, in all of the aforementioned studies, with the exception of Kingston & Pappas (2009), shape group differentiation was not necessarily completely satisfied. As indicated previously, diatom shape is a continuous variable with respect to vegetative size reduction. If diatom shape variation is minimal, Legendre polynomial and multivariate analysis may not be sensitive enough to extract distinct shape groups. (See the sections “Outline shape, orthogonal polynomials and their properties, and regression coefficients,” and “Choice of method and issues to consider” for more details).

For studies using the discrete Fourier transform, results were produced to hypothesize distinct shape groups, while highlighting the necessity to apply novel methods to extract information on fine differences in variation (e.g., Pappas 2006). In a study of *Tabellaria*, Fourier descriptors were used to delineate eight shape groups (Mou & Stoermer 1992). For *Asterionella*, seven shape groups were found (Pappas & Stoermer 2001). However, in both of these cases, a high degree of shape variability indicated the continuous nature of valve shapes produced by each taxon via size reduction. With diatom variation as a continuous variable, both minimal shape variation (mentioned above) and high degree of shape variation may present difficulties in extracting shape groups.

More recently, automated diatom identification has reappeared in a more encompassing form with the advent of the Automatic Diatom Identification and Classification (ADIAC) project (du Buf & Bayer 2002). A group of researchers devised a number of unsupervised contour and feature extraction and identification systems using the power of computers, mathematical modeling, microscopy, and information about morphology, taxonomy and ecology of diatoms.

For contour extraction (Fischer et al. 2002), segmenting gray level images of diatoms is used (Loke & du Buf 2002). Edge detection (e.g., Canny 1986) or thresholding (e.g., Sahoo et al. 1988) is used to differentiate diatoms in a given field by detecting the difference among diatom outlines (as closed contours) and other unstructured background gray level matter. Feature extraction (Fischer & Bunke 2002), as texture descriptors, is also used in making diatom identifications. The process is automated with respect to slide scanning autofocusing and microscopy, and decision trees or forests and other means of classification are used to achieve identifications. Various methods are used in the segmenting, extraction and identification processes, including Gabor functions (Santos & du Buf 2002), dynamic ellipse fitting (Ciobanu & du Buf 2002), mathematical morphology (Wilkinson et al. 2002), and a combination of such methods. Results

from the system were compared with those produced by experts in diatom taxonomy. In addition, comparisons of results were made to those using Fourier analysis (e.g., Fischer & Bunke 2002) or Legendre polynomials (e.g., Ciobanu & du Buf 2002, Mann et al. 2004), as these were methods previously used in diatom research. Subsequently, other diatom recognition studies involve, for example, extraction and grouping of contour segments of diatom shape (Loke et al. 2002), using segmentation and feature extraction to distinguish between diatoms and debris in water samples (Forero-Vargas et al. 2003), building curvature scale space and extracting features using cluster analysis (Jalba et al. 2005), and using Fourier shape descriptors, texture analysis using the fast Fourier transform and dimensionality-reduced probabilistic principle curves (Hicks et al. 2006). Overall, a large contribution was made by the researchers involved with ADIAC to the compilation of mathematical methods previously applied in diatom identifications.

Recently, Beszteri et al. (2005) and Potapova & Hamilton (2007) used landmark-based morphometrics on *Cyclotella meneghiniana* Kützing and related species and the *Achnantheidium minutissimum* (Kützing) Czarnecki species complex, respectively. Subsequently, Falasco et al. (2009), Novais et al. (2009), and Vesela et al. (2009) used these initial studies to apply landmark-based methods to diatoms. Until this time, landmark-based morphometrics were not used on diatoms since questions were raised about the applicability (Mou & Stoermer 1992). It should be noted that some practitioners of geometric morphometric methods have recognized that landmark-based methods and outline methods are not incompatible when used in combination for analysis. Combination landmark-based and outline methods include sliding semilandmarks (e.g., Bookstein 1996), edgels (rotation of direction through a landmark to acquire information about curvature) (Bookstein & Green 1993, Little & Mardia 1996), and creases (global extrema as landmarks and local outline points with respect to degree of smoothing to obtain a singularity) (Bookstein 2000). In spite of this, there are still a number of issues to be dealt with in deciding whether landmark-based methods can actually be used to analyze diatoms to achieve reliable results. This will be discussed below.

Using landmark-based methods in diatom research

Two recent papers using landmark-based methods in diatom research illustrate some of the challenges involved in using such methods. Number of specimens used, biological and geometric homology, sampling all members of a size reduction series, the number of landmarks used, the location of landmarks, measurement error, and choice of a reference are among the issues that require careful consideration. All of these matters will affect the validity of the outcome from landmark-based analysis.

Beszteri et al. (2005) used both conventional and landmark-based morphometric methods in their study. (By conventional morphometrics, Beszteri et al. (2005) mean traditional morphometrics or simple valve measurements). Using the methods of thin plate splines and partial warps in their application of landmark-based morphometrics, they sought to differentiate *Cyclotella meneghiniana* from *C. scaldensis* Muylaert & Sabbe and other presumably different morphological variants. The landmarks they identified involved the rimoportulae on the valve margin and the centroid of the valve face. They did not include all specimens that were used in their traditional analysis since some of them did not qualify within their own specifications.

If all specimens are not included in an analysis of shape the results may be biased (Adams et al. 2004). In a study of shape variation limits in orders of mammals, Marcus et al. (2000) found that such variation among skulls is low. However, they also acknowledged that their results were misleading since some landmarks could not be used because they did not exist on some of the taxa used in the study. The limitations of either not using all specimens or not using all landmarks distort the outcome of such studies (Adams et al. 2004).

Historically, homology has acquired many meanings. In biology, homology is a concept used to signify similarity between anatomical structures as a result of some original structure in a shared common ancestor (e.g., Wiley 1981), structures that arose from the same developmental

origin (e.g., Wagner 1989), structures defined by a specific function (e.g., Cracraft 1967), or structures that originated genetically with regard to DNA sequences, proteins or chromosomes (e.g., Ghiselin 1969). According to Rohlf (1998), homology with regard to morphometrics means “geometric” structures between two organisms that have a consistent biological or biomechanical meaning that are represented by points, curves, and subsequently, by multivariate descriptors (also see Slice et al. 1996). Again, landmarks as homologies are biologically meaningful only when they are based on phylogeny or development (Cartmill 1994) or genetics or functionality. Determining biological homology was traditionally based on whole structures, not individual geometric points (MacLeod 1999). However, unlike biological homologies, landmarks might also be identified that are only meaningful in a geometric (= mathematical) sense (e.g., pseudolandmarks), or landmarks might represent more than one region of a form (Rohlf 1988). The qualifications to meet regarding the various definitions of landmarks determine in great measure how the results of landmark-based analyses are interpreted. Ideally, biological homology is the goal of landmark-based analyses.

Diatom size reduction involves changes in the characters (as homologous structures) on the valve face, and detecting these changes depends on whether all stages are present and are sampled. Unlike their traditional morphometric treatment, Beszteri et al. (2005) did not specify whether complete size reduction series were used in landmark-based analysis. Depending on what stage in a size reduction series is present, it is also important to specify which characters on the diatom valve face that are chosen as landmarks are invariant in order to yield biologically meaningful results. Ontogenetic correspondence is particularly important in diatoms.

As stated earlier, function is important with regard to homology (e.g., Mou & Stoermer). If the function of the structure (= character) is in question, then assigning landmarks to the structure for partial and relative warp analysis is unwarranted. If the function of the rimoportula is known sufficiently well, and landmark-based methods are used to its determine shape variation interspecifically, then the results from analysis will be biologically meaningful.

In general, choosing landmarks is not an easy task. The relative importance of the points chosen is ultimately a subjective matter in the purview of the researcher doing the analysis. However, choice of landmarks, including the centroid, is only valid if landmarks that are widely dispersed on the valve face are chosen (Bookstein 1991). The centroid on a diatom valve is meaningful only in regard to its shape perimeter, and shape perimeter was not the focus of Beszteri et al.’s (2005) study. Instead, Beszteri et al. (2005) chose the centroid, which is distant from the landmarks chosen that are clustered close together on or near a rimoportula. Deformation between two points chosen far away from each other will be highly distorted in contrast to two points close together, which move approximately in tandem (Bookstein 1986). That is, points chosen in close proximity that move in approximate tandem will not measure shape change (Bookstein 1986).

Choice of landmarks also entails other caveats of which researchers must be cognizant. Landmarks can induce two forms of measurement error. One is with regard to digitizing coordinates; the other is with respect to intrapopulation variation in shape (Goodall 1991). That is, measurement errors may be found in geometric features or, alternatively, in the size-shape spaces generated (Bookstein 1986). In addition, choosing landmarks means that highly curved parts may be missed in the analysis (Bookstein 1991). As a result, curving of form in between landmarks is not recovered in landmark-based Procrustean analysis. In Beszteri et al.’s (2005) study, the curved part of the rimoportula is not sampled between landmarks, so the curvature of the rimoportula was not analyzed.

Sampling relevant locations of an organism to recover morphological structures or shape is necessary. If measuring shape change among a group of organisms is the goal of the study, then inadequate choice of landmark locations will mean that shape changes will not be measured and the results will be misleading (Lestrel 2000). If landmarks chosen are not independent of one another, they are not legitimate points to be used in Procrustean analysis. Moreover, shape variation patterns need to be tested with respect to the population mean shape to be sure that such

patterns are not mathematical artifacts of superimposition, since it is possible for superimposition to induce a covariance structure on the landmarks (Adams et al. 2004). The number of landmarks chosen is also a matter of concern, since this will influence the outcome of the analysis (Adams et al. 2004). The limitations of landmark-based methods do not necessarily extract morphologically relevant information in the best or most optimal way (Lestrel 2000).

Once relevant landmark locations are determined, landmark-based methods involve multivariate analysis by extraction of principal and partial warps (Slice et al. 1996) and the calculation of relative warps (Bookstein 1991). Partial warps are nonaffine transformations of localized bending energy. Depending on the value of the eigenvalues, small or large scale (local or global) deformation is represented. That is, the first principal and partial warps (= small eigenvalues) are associated with large scale deformation, while the last principal and partial warps (= large eigenvalues) are associated with small scale deformation and represent landmarks that are close together. With respect to principal and partial warps, relative warps are used when a random sample of shapes exists, and this alternative is similar to using PCA with a Euclidean metric (Dryden & Mardia 1998).

Partial warp scores form a continuous variable (Bookstein 1991, Rohlf 1998), and these scores along with relative warp scores must be used cautiously. Beszteri et al. (2005) regressed relative warp scores on centroid size and valve diameter in order to account for allometric changes. Valve diameter relates to the overall size of a specimen and not necessarily to the size of the rimoportulae. If valve diameter is confounded by shape because of developmental considerations, then it is not necessarily independent of size, and this will influence regression of partial warp scores on valve diameter. Beszteri et al. (2005) did not report correlation coefficients from regressing partial warp scores on centroid size and valve diameter.

There has been much debate on the use of partial and relative warp scores in other capacities with regard to what exactly they measure (e.g., Rohlf 1998, Adams & Rosenberg 1998). Individual partial as well as relative warp scores are not biologically interpretable as single measures (Adams & Rosenberg 1998). They are the eigenvectors or axes and are only biologically meaningful when considered over the sum of all axes that is tangent to the total shape space (Bookstein 1991, Rohlf 1996). Eigenvalues for each relative warp and the trace indicate the degree of variation explained by the analysis. Beszteri et al. (2005) did not report eigenvalues from their relative warp analysis.

The principle warp axes are based on the reference form chosen. These axes are sensitive to the choice of a reference (Rohlf 1996). Beszteri et al. (2005) chose to use the mean shape-determining landmarks as a reference, and this included the central part of the valve. Overall, in a diatom size reduction series, the mean shape may not be representative, and interspecifically, comparison of taxa may be difficult. If overall shape is to be analyzed, perhaps a median would be a better reference, but this would need to be tested.

Using traditional morphometrics, Beszteri et al. (2005) study produced separation of shape variants among taxa. Using landmark-based methods, Beszteri et al. (2005) found it difficult to interpret differences between *Cyclotella scaldensis* morphs. They also found that values from cultures had more shape variation using landmark-based methods than using traditional methods. As Beszteri et al. (2005) have indicated, landmark-based methods in diatom research are not applicable in cases where the “varying number of most structures found on diatom frustules is ... a practical problem”. As a preliminary study, Beszteri et al. (2005) have shown that application of landmark-based methods in diatom research has many caveats to overcome in order to produce reliable results. Unlike shape outline analysis; there are particular problems that must be addressed when attempting to use landmark-based methods in diatom research.

Potapova & Hamilton (2007) used sliding “landmarks” (Bookstein 1996, 1997) in their study of *Achnanthydium minutissimum* and related taxa in order to circumvent the problem of identifying biologically homologous points. This modification of the original usage of landmarks was devised since using landmarks, as specified previously, is not always applicable for a number of reasons. To reiterate, landmarks may not be identifiable for many scientifically important

organisms or parts thereof (Bookstein 1996, 1997), such as areas of high curvature (Perez et al. 2006). In addition, location of landmarks may be rendered uncertain because of insufficient confidence that they are applicable on a one-to-one basis among a group of forms or across a given range of shape variation (e.g., Perez et al. 2006). That is, there is uncertainty in locating landmarks on curved areas when few to no landmarks are present elsewhere (Bookstein 1996, 1997, Perez et al. 2006). Finally, landmarks form a discrete data set that does not extend to explanations about continuous data that are biologically important, such as surface areas and volumes (Bookstein 1996, 1997), if this is a goal of the research being conducted.

Sliding “landmarks” are actually semilandmarks (e.g., Perez et al. 2006) or quasilandmarks (e.g., Bookstein 1997) whereby points on a curved outline are slid on one form in order to match them with points on a reference form (Green 1995, Bookstein 1996, 1997). To match forms means removing tangential variation by minimizing bending energy (Bookstein 1997) or Procrustes distance (Bookstein et al. 2002) via thin-plate splines (Bookstein 1996). Semilandmarks may be combined with actual landmarks using this technique of matching contours via sliding points. The rationale is that contours, not the points, should be homologous from shape to shape. In their study, Potapova & Hamilton (2007) identified the points they chose for study as landmarks, however, this is an error since only semilandmarks (or quasilandmarks) can be used with the sliding technique. For the purposes of discussion, we will use the term ‘semilandmark’ in place of their designation.

Initially, 16 semilandmarks were chosen around the periphery of the diatom valve (Potapova & Hamilton 2007). Asymmetry was corrected for by averaging x , y coordinates from mirror images of valve faces (= reflection), resulting in an object symmetry (e.g., Mardia et al. 2000) for each valve face. From this, one-quarter of the valve face semilandmarks were used in Procrustean analysis, as was the centroid of the valve face (Potapova & Hamilton 2007). That is, Potapova & Hamilton (2007) used 5 semilandmarks and 1 landmark for a total of 6 points.

Semilandmarks are sensitive to location (Adams et al. 2004). In addition, the number of semilandmarks used will affect the results. If more semilandmarks than landmarks are chosen, the semilandmarks will outweigh the landmarks and skew the results toward curvature (Perez et al. 2006). If fewer semilandmarks are used, then there will not be enough value about outline curvature extracted from the analysis so that shape information is recovered. There is also the possibility that too few semilandmarks may be chosen because of the nature of the outline shape, so that recovering the outline does not completely occur. By contrast, with shape outline methods, the only requirement is not to choose so many points as to cause overfitting (e.g., Pappas et al. 2001).

Overall, use of semilandmarks and outline methods may not be very different from one another in the sense that they are used to interpret biological homology in the same way (MacLeod 1999, Sheets et al. 2004). The similarity between outline and semilandmark methods is evident when examining the way in which coordinates are selected on the diatom valve outline. In outline methods using Legendre polynomials, coordinates chosen must be equidistant from each other. When using the discrete Fourier transform and the method of arc lengths and tangent angles, coordinates are chosen by “sliding” them into position relative to other specimens regardless of size. It is not necessary to use equidistant coordinates, but sometimes this occurs by happenstance. Both outline and semilandmark methods rely on homologous outline segments for use in analysis.

For their reference shape, Potapova & Hamilton (2007) chose the average of the average coordinates of reflections for each image and called this a ‘consensus configuration’. That is, they used a “consensus” figure for each diatom valve face, and then took the average of all these “consensus” figures to use as a reference shape. The mean shape may not be representative with respect to size reduction series. Potapova & Hamilton (2007) did not use complete size reduction series, and they used unknown stages with respect to life cycle. They tried to eliminate allometric variation in their data by regressing partial warp scores versus valve length, and then use the residuals as shape variables. That is, they regressed a shape dependent variable uncorrected

for size (partial warp scores) on a size dependent variable (length), with the left over variation used as a size-independent shape variable (residuals). However, partial warp residuals are not independent variables, and they are used for investigating shape variability, not size-independent shape variability ([Dryden & Mardia 1998](#)).

Subsequently, [Falasco et al. \(2009\)](#), [Novais et al. \(2009\)](#), and [Vesela et al. \(2009\)](#) in their studies of the *Sellaphora stroemii* species complex, *Gomphonema rosenstockianum* – *G. tergestinum* species complex, and *Navicula cryptocephala* – *N. trivialis* species complex, respectively, based their methodologies on the previously discussed papers. In their studies there was a lack of understanding of the difference between landmarks and semilandmarks, with [Falasco et al. \(2009\)](#) and [Novais et al. \(2009\)](#) using them for morphometric analysis, but not mentioning this in the text. In [Vesela et al. \(2009\)](#), there was misidentification of semilandmarks as landmarks. As with the previous papers, the aforementioned problems with using geometric morphometrics in diatom research were not addressed. Our attention will be focused next on shape outline methods since they have been tested over time. In addition, we are interested in examining dimensionality reduction methods that are used to analyze shape descriptors and how these methods are related. We are primarily interested in shape analysis where homologous curves and biologically meaningful interpretation of results is the desired outcome in shape space, rather than those methods that produce only geometrically meaningful results.

Using outline morphometric methods in diatom research

In general, as has been already stated, there are two areas of study with regard to outline methods used in diatom research, namely, shape analysis and pattern recognition. The differences between shape analysis and pattern recognition can be summarized as follows: pattern recognition is the detecting of differences in shape and pattern among specimens for taxonomic decision-making and binning of specimens, while shape analysis is used to find such differences in shape and pattern that can be interpreted as biologically meaningful in a morphospace.

Both shape analysis and pattern recognition have been shown to be useful in research involving diatom taxonomy and classification (e.g., [Stoermer & Ladewski 1982](#), [du Buf & Bayer 2002](#)) schemes, where such schemes have been used in evolutionary, ecological, or other studies. With many questions remaining in diatom taxonomy, development, ecology, among other related fields, caution should be exercised in deciding which morphometric methods are appropriate for use. Choice is dependent upon individual research questions, and the researcher's purposes and level of understanding dictate which methods should be used. Regardless, a thorough understanding of the methods chosen for use will promote success in the results obtained and confidence by the readership in the legitimacy of those results.

With this in mind concerning geometric morphometrics, biologically meaningful results from shape analysis are the most desirable goal in using such methods. Applying mathematical analysis to further understanding of the biology of diatoms and also as an aid in decision-making in diatom taxonomic and related problems should be the rationale for using morphometrics. However, there will be times when biological evidence is scant or inaccurate so that recourse to mathematical constructs might be the only way to hypothesize, for example, about potential connections among specimens in diatom taxonomic studies. At the very least, applying mathematical methods, such as shape analysis should be used with great care and understanding.

In the following subsections, we will focus on the use of shape analysis and resultant shape morphospaces to aid in understanding the methods employed. Some details are provided about diatom shape descriptions, basic measurements, orthogonal polynomials and their properties and usage in regression, issues to consider, how to treat shape descriptor data, and multivariate methods used to depict shape space.

Diatom valve shape as a morphological character

Categorizing diatoms geometrically is a starting point to determine which methods are appropriate for shape analysis. Many centric diatoms are circular, while others are not. According to Barber & Haworth (1981), shape descriptions may be ascribed to diatom valves, and for shape analysis these descriptions may be useful in matching shape description as a category with shape analytical method as well as interpreting such descriptions with respect to method used to calculate n -dimensional shape space.

As an overview, some diatom valve shapes, generally pennate diatoms, may be categorized as mostly oblong, which include oval and rectangular shapes. Those that are mostly oval go by the shape descriptions of “elliptic”, “narrow elliptic”, or “ovate”, while rectangular shapes may include “acicular”, “lanceolate”, or “linear” as descriptive attributions. Also included in the rectangular-like subgroup are triangular, rhomboidal and trapezoidal shaped valves with straight sides (Barber & Haworth 1981). All of these shapes include sections of the valve that are straight for most of the outline where the change in shape is zero.

All other diatom valve shapes may be described as asymmetrical or highly curved. Subgroups of shapes that are asymmetrical with respect to the apical or transapical axis include those that are “clavate”, “panduriform”, “reniform”, “crescentic”, or “arcuate”. Highly curved shapes include “biundulate”, “triundulate”, or “ n -undulate” as well as “cruciform”, “sigmoid” and “polygonal” shapes without straight sides (Barber and Haworth 1981).

Further subdivisions of categories can be made with regard to shape of the ends of pennate diatom valves. Descriptive terms such as “acute”, “apiculate”, “rostrate”, “capitate”, “subcapitate”, “rounded”, “cuneate”, or “spatulate” can be attributed to diatom valve ends (Barber & Haworth 1981) where the main part of the valve shape is either oblong/rectangular (straight sides) or highly curved/asymmetrical.

All of these categories may be modeled mathematically. In the following, we only look at methods of shape analysis rather than pattern recognition. We want to explore why shape analysis methods are useful in extracting biologically relevant information in diatom studies, why these methods work in the context of a morphospace, and why these methods may not work in all cases. In addition, we explore the possibilities for new morphometric methods that should be used chiefly to extract biologically meaningful information for use in diatom studies.

Basic measurements

When using any method of outline shape analysis, it must be remembered that diatom valves are three-dimensional objects, hence any representation of valve shape is dependent upon the choice of measurement plane. For species with complex forms this may not be a simple choice, particularly if the light microscope is used for measurement. For most species, the easiest and probably most reliable solution is to use the lower edge of the valve mantle as the plane of measurement. This provides a shape plane of focus and is more easily defined than other potential reference points of the valve. For specimens such as *Achnantheidium* that are highly curved in the valvar plane, this method is not satisfactory for recovery of whole valve outlines, but may be used for recovery of partial outlines in some specimens. Needless to say, a great deal of care is needed to provide accurate measurements. Very slight errors in measurement may accumulate and provide confusing results.

Outline shape, orthogonal polynomials and their properties, and regression coefficients

For shape analysis of diatom valve outline, orthogonal polynomials such as Legendre polynomials (e.g., Stoermer & Ladewski 1982) and the discrete Fourier transform have been used (e.g., Mou & Stoermer 1992). Starting with a finite number and position of coordinates on the diatom valve outline, orthogonal polynomial regression is used by determining the curved segments of the outline, for example, as change in width function (Legendre polynomials) or

change in arc lengths and tangent angles (discrete Fourier transform). Orthogonal polynomial regression is a method of fitting a least-squares curve around an outline. Orthogonal polynomials have properties that define their suitability in analyzing valve shape with respect to the shape categories specified. From x, y coordinates of a diatom valve outline, coefficients from orthogonal polynomial regression are used as numerical shape descriptors.

Of all the orthogonal polynomials, Legendre, Chebeshev (of the first and second kind), Gegenbauer, and Zernike polynomials are special cases of the Jacobi polynomials (Abramowitz & Stegun 1972). Other orthogonal polynomials commonly known are Hermite and Laguerre polynomials (Abramowitz & Stegun 1972). Legendre polynomials are the simplest of the orthogonal polynomials because their weight function is equal to 1, unlike all the other orthogonal polynomials (Abramowitz & Stegun 1972). For n -Legendre polynomials, $P(x)$, that are recursive (Szokefalvi-Nagy 1965), and n -Legendre coefficients, c , that are used as shape descriptors, expansion of the width function, $W(x)$, is

$$W(x) = \sum_1^N c_n P_n(x) \quad (1)$$

(Stoermer & Ladewski 1982, Pappas & Stoermer 2003). The width function is a linear combination of Legendre polynomials of degree n in x , and each Legendre coefficient, c_n , is an independent shape descriptor because of orthogonality. Legendre polynomials form a complete orthogonal system based on the interval $[-1, 1]$ and a weight function (ρ) of 1. Expansion of the width function represents the expression of Legendre polynomials as a series.

The generalized orthogonality relation may be stated as

$$\int_j^k f_n(x) f_m(x) \rho_n(x) dx = c_n \delta_{mn} \quad (2)$$

where, in the case of Legendre polynomials, $j = -1$ and $k = 1$ are boundaries of the interval for the width function, $f(x) \equiv W(x)$, with weight function $\rho = 1$, and δ is Kronecker delta. Because of orthogonality and expansion of the width function, Legendre polynomials as a series are mathematically related to Fourier series. That is, Legendre and Fourier expansions are orthogonal expansions. Expansion of the Fourier series may result in the Fourier transform (Edwards 1967), and it is coefficients from this that are used as shape descriptors. So, expansion of the Fourier series takes the form of

$$\Phi(x) = \sum_1^N c_n \varphi_n(x) \quad (3)$$

where $\varphi_n(x)$ are the Fourier functions of $\cos(nx)$ and $\sin(nx)$, and in the generalized orthogonality relation (Eq. (1)), $f(x) \equiv \Phi(x)$, c is the n th coefficient, $[j, k]$ is the interval $[-\pi, \pi]$, and the weight function, ρ , is equal to 1 (Weinberger 1995).

For orthogonal polynomial regression using Fourier functions, we see that

$$\int_j^k \left[f(x) - \sum_n^N c_n \varphi_n(x) \right]^2 \rho(x) dx \quad (4)$$

becomes

$$\int_j^k f(x)^2 \rho(x) dx - \sum_n^N c_n^2 \int_j^k \varphi_n(x)^2 \rho(x) dx \quad (5)$$

and

$$\sum_n^N c_n^2 \int_j^k \varphi_n(x) \rho(x) dx = \int_j^k f(x)^2 \rho(x) dx \quad (6)$$

or for Legendre polynomials,

$$\sum_n^N c_n^2 \int_j^k P_n(x) \rho(x) dx = \int_j^k f(x)^2 \rho(x) dx \quad (7)$$

We want to minimize the difference between $f(x)$ and Legendre polynomial or Fourier function by the choice of coefficients, c . That is,

$$\|f(x) - \varphi_n(x)\|^2 = \left\{ \int_j^k \left[f(x) - \sum_n^N c_n \varphi_n(x) \right]^2 \rho(x) dx \right\}^{1/2} \quad (8)$$

where the norm is a measure of the deviation of $\varphi_n(x)$ from $f(x)$ (Brown & Churchill 2001). From this, it should be evident that Legendre polynomials and Fourier functions as orthogonal functions using least-squares regression, can produce orthogonal coefficients in a similar fashion that are the basis of shape descriptors in shape analysis.

Coefficients from a linear combination of orthogonal polynomials are used as shape descriptors. In some cases, these coefficients may be directly interpretable as numerical representations of morphological features of diatom valve shapes. That is, numerical shape descriptors may correspond directly to morphological features of a diatom valve outline. This is a desirable property in that some biologically meaningful information should be extractable from a numerical process of analyzing diatom shape.

Such a desirable property has been demonstrated by using Legendre polynomials (Stoermer & Ladewski 1982) but is not readily possible with respect to Fourier functions (e.g., Mou & Stoermer 1992, however, see Ehrlich et al 1983). Matching geometric constructs to shape morphology and descriptions may be done with each successive Legendre polynomial. Using this, a model is built up from one Legendre polynomial to the next whereby there is an associated physical description of morphology of the valve sides, middle and ends at each step in the process.

Overall, the discrete Fourier transform is useful in modeling shape for any diatom outline despite lack of matching Fourier coefficients to morphological features. Legendre polynomials may be used when diatom valve outline is highly or continuously curved, or at least, contains closely spaced curved portions. If diatom valve outline is less curved and more elliptical or rectangular or has long spans that are linear, the Fourier transform would be more suitable. For example, *Tabellaria* has relatively long, straight segments interrupted by areas of extreme curvature in the middle and at the ends of the valve. In the case of *T. flocculosa* (Theriot & Ladewski 1987), valve outline was not recovered as well using Legendre polynomials. By

contrast, a subsequent study of *Tabellaria* shape was modeled using the discrete Fourier transform and outline shape was more successfully recovered (Mou & Stoermer 1992). Choice of a particular function depends not only on the curvature of the valve outline, but also the particular mathematical properties of the function selected for use. Although using the discrete Fourier transform is not necessarily a method to directly match shape coefficients to physical morphological descriptions of the diatom valve, it is suitable in calculating shape descriptors that may be used in subsequent analysis to determine shape groups. Whether these shape groups are species groups is another matter. This method may be one of the only ways to represent diatom valve outline numerically, especially for character-poor taxa having shapes similar in curvature to *Tabellaria* and *Asterionella*.

In a recent study of the *Sellaphora pupula* (Kützing) Mereschovsky species complex, Mann et al. (2004) compared using Legendre polynomials to morphometric constructs using contour segment analysis based on the Fourier transform. From their results, Legendre coefficients ordinated in shape space using PCA were dispersed as three groups, while results from individual plots of maxima or minima contour points or curvatures were depicted as six groups. Mann et al. (2004) did not use multivariate analysis or cross-validate shape groups from results of contour segment analysis. Their interpretation of results is a comparison of apples to oranges since a multivariate method of shape variation (PCA with Legendre coefficients) was compared to their univariate one. Moreover, as stated previously, Legendre polynomials and coefficients, when ordinated in shape space, are directly interpretable and matchable to physical morphological features of valve shape, while diatom shape analysis based on the Fourier transform has not been shown to be interpretable as such (e.g., Mou & Stoermer 1992, Pappas & Stoermer 2001). It may be that Legendre polynomials are not appropriate for use since the outlines of some specimens in the *S. pupula* species complex have curved sections interrupted by more linear segments, and this was not tested. Most importantly with regard to multivariate methods, the method of dimensionality reduction has more to do with differentiating shape groups than the method used to obtain shape descriptors. The specifications of the methods to extract shape descriptors, namely Legendre polynomials, are not the means to classify shapes. That is, the resultant shape coefficients from Legendre polynomials are the data, while multivariate methods are used for classification and shape group separation, and therefore it is these methods that have the most influence in what kind of shape group separation occurs and how well separation of shape groups is defined. Choice of the multivariate method used to analyze shape data is primarily responsible for depiction of shape group separation, and others have corroborated this finding (e.g., Sheets et al. 2006).

Shape descriptors and data transformation

Perhaps the issue is the actual degree of difference among the shapes being considered for shape group determination. If the difference in shape is small, then it may be that whatever the multivariate method used will not be sensitive enough to reveal separate shape groups readily, even in the case where the number of species groups is known *a priori*. That is, the distribution of the data used as shape descriptors may be narrowly defined. This situation may require data transformation in order to disperse the scores in an ordination.

In a study of *Asterionella* using Fourier coefficients (Pappas 2000, Pappas & Stoermer 2001), 22 mean-corrected amplitudes, \mathbf{A} , were arranged in an $m \times n$ matrix of m rows (specimens) by n columns. To improve symmetry about the grand mean, data transformation was used (Noy-Meir 1973, Noy-Meir et al. 1975). That is, centering on rows (specimens) and standardizing on the row (specimen) norm was used to move the centroid to the origin and rescale the variance around the origin (Noy-Meir et al. 1975, Jongman et al. 1995). Centering on rows (specimens) ensured that the contribution of each specimen was proportional to its variance (Noy-Meir et al. 1975). Standardization by row (specimen) norm, where the mean equals zero and variance equals one, produced equal contribution of all specimens. In this way, no one specimen would unduly

influence the ordination. In addition, square root transformation was used to further disperse the data around the grand mean since between-shape variance for all *Asterionella* was within a narrow range (Pappas 2000, Pappas & Stoermer 2001).

Multivariate methods and the analysis of shape variation

Variation in shape coefficients is depicted in an ordination as a result of multivariate statistical analysis. This is typically the way in which to analyze all valve shapes with respect to each other. Multivariate shape space has properties that make it useful in shape analysis and discerning differences in shape variation on an n -dimensional basis.

Multivariate shape space may be characterized as the result of the singular value decomposition (e.g., Green & Carroll 1978). Using the aforementioned example of the matrix of mean-corrected amplitudes as *Asterionella* shape descriptors, \mathbf{A} , the singular value decomposition of \mathbf{A} may be stated as $\mathbf{A} = \mathbf{PDQ}'$ where \mathbf{P} is the $m \times r$ matrix of eigenvectors of the major product moment \mathbf{AA}' . The $n \times r$ rotation matrix, \mathbf{Q} , is the matrix of eigenvectors of the minor product moment $\mathbf{A}'\mathbf{A}$. The diagonal matrix, \mathbf{D} , has positive rank-ordered eigenvalues with off-diagonal elements equal to zero.

Although shape space is calculated to be multidimensional, the idea in using multivariate methods is to find the fewest number of independent eigenvectors that represent the greatest amount of shape variation. Underlying eigenvectors extracted are orthogonal and, if normalized to unit vectors as stated above, orthonormal. The result is that shape space represents a new coordinate system of shape descriptors where an ordination is a picture of the variance-covariance matrix of all shape values in the lowest dimensional space possible to explain the maximal amount of shape variation. Chiefly, PCA and discriminant analysis (DA) (including canonical variates analysis (e.g., Goldman et al. 1990) and multivariate analysis of variance (MANOVA) (e.g., Steinman & Ladewski 1987) have been used to calculate shape variation, and we focus on these methods for use in calculating n -dimensional shape space.

Principal components and discriminant analysis

Initially, PCA is used as the dimensionality reducing method for shape descriptors. For the most part, PCA is used initially to depict total shape variation. In general, most of the shape variation in diatoms is related to size, and this allometric feature is revealed in PCA where ordination depicts subsequent parts of total shape variation on a dimension-by-dimension basis. In this way, PCA can account for shape variation numerically and concisely. For PCA, to obtain \mathbf{Z} principal component axes, $\mathbf{A} = \mathbf{PDQ}'$ is postmultiplied by \mathbf{Q} so that $\mathbf{AQ} = \mathbf{PD} = \mathbf{Z}$. In addition, the principal components are rescaled so that their variances are equal to one so that standardized unit variance component scores are calculated as $\mathbf{Z}_s = (\mathbf{m} - 1)^{-1} \mathbf{AQ}\Delta^{-1}$ (Green & Carroll 1978). From this result, component scores are dispersed in shape space so that additional analysis could be performed on the total shape variation gradient (Pappas 2000, Pappas & Stoermer 2001).

In addition, it is possible, though not necessarily the case, that while total shape variation may be depicted in a PCA ordination, shape groups may be evident when separation between clusters of PCA scores occurs. Separation of shape groups in PCA is influenced by other factors such as sample size and the extent to which the entire spectrum of shape variation was sampled. Moreover, it is not necessarily the case that differences in shape groups should be evident concomitant with the eigenvectors representing most of the variation or with sequential eigenvectors plotted as two-dimensional ordinations. Small eigenvalues or non-sequential eigenvectors may be sources of explanation of differences among shape groups (e.g., Stoermer & Ladewski 1982).

While PCA is useful for total shape variation, DA (e.g., Cooley & Lohnes 1971, Green & Carroll 1978) is the way in which to identify actual shape groups, and perhaps, even corroborate such shape groups as distinct species. Moreover, DA may be viewed as a way to cross-validate shape groups that have been initially characterized by PCA. This point is critical in understanding why choice of multivariate method is essential in providing valid, credible results.

If the number of potential species groups is known prior to analysis, DA is the gold standard for use. When the number of groups is not known, other methods may be used to approximate a way to cross-validate what is hypothetically thought to be, species groups. However, a great deal of understanding and effort is necessary to show that other methods may produce results that have a similar prospect of cross-validation that DA has when it comes to validating group designations.

DA is based on the generalized or Mahalanobis distance (Mahalanobis 1948). This distance takes into account the correlation among variables and is used to measure the distance between the center of the distribution of all shape descriptors and each individual shape descriptor (Manly 2005). Then, each shape descriptor is a member of the group to which it is closest. The significance of the difference among group centroids is tested by Wilks' lambda, Λ , as

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} \quad (9)$$

where \mathbf{T} is total shape variance, \mathbf{W} is within-group variance, and $\mathbf{T} - \mathbf{W} = \mathbf{B}$ (Manly 2005), where \mathbf{B} is between-group variance. The smaller Wilks' lambda is, the more significant is the difference among group centroids. The idea with DA is to minimize the variance among a group's members, while maximizing the variance among all groups.

As in PCA, DA is based on the singular value decomposition. Briefly, a matrix of canonical eigenvectors, \mathbf{Z} , is calculated as

$$\mathbf{Z}_s = \mathbf{A}\mathbf{K} \quad (10)$$

where \mathbf{A} is the $m \times n$ matrix of shape coefficients and \mathbf{K} is the matrix of discriminant weights (Green & Carroll 1978). \mathbf{Z} are the canonical eigenvectors to be ordinated in shape space.

Like PCA, DA is dependent on sample size and whether a sufficient representation of shape variation is sampled influences the outcome of analysis. In turn, the decision about number of specimens to use is dependent on what is available, what the quality of the specimens are, what research question is to be answered or what the hypothesis is to be tested, what kind of instrumentation, equipment and computer power is available for analysis, and what knowledge-base is available in terms of understanding and implementing the methods of data analysis and interpretation of results. All these factors need to be considered to develop a shape analysis study utilizing PCA and DA to get meaningful results.

Choice of method and issues to consider

Even with data transformation, and choice of the appropriate multivariate method, results may not be as clear as one would hope. In particular, it may be that Legendre polynomials or the discrete Fourier transform may not adequately model shape coefficients that are ordinated in clearly distinct groups. Other factors are involved and other techniques may be used to further treat the data. In an ordination, if the density of shape scores is such that overlap of potential shape groups occurs, a number of methods may be used to divide the shape gradient into potential groups. Sectioning by ellipses representing approximate normal distributions may be used (e.g., Mou & Stoermer 1992). However, there is no way to test the assumption of normality since the sectioning is somewhat arbitrary. Partial classification based on other morphological, ecological, geographical, or other data may be used (e.g., Pappas & Stoermer 2001, Pappas 2006). Using additional empirical evidence is helpful in lending credence to using partial classification methods. However, in no way does this guarantee clearly defined groups that may be identified as "species."

How much evidence is sufficient for a shape group to be named a species group? Ideally, a multitude of biological evidence should already exist to prompt the creation of a hypothesis

about a group of taxa. A generally accepted definition of “species” is necessary in order to make judgements on the efficacy of the application of shape analysis. Within that definition, it is necessary to determine how much morphological change is equal to species differences, given that diatom taxonomy is still largely based on morphology and information based on cytology, ecology, physiology, among other things, is unavailable for many taxa.

Shape analysis may become important in a number of situations. One such situation is when some biological information is missing, unreliable, or historically incorrect, but many specimens are available for study. Shape analysis may be used to develop hypotheses that could be tested empirically when evidence becomes available. Another situation involves difficulties in visually identifying taxa. This is a well-known problem in diatom studies, especially with small taxa or specimens that are size reduced as a result of the vegetative life cycle. Shape analysis may be used to numerically differentiate specimens that are visually similar. Finally, as mentioned above in the context of discriminant analysis, shape analysis may be used to cross-validate species groups that have been determined by biological evidence alone.

Shape descriptors, shape space and biologically meaningful interpretation

Past shape analysis studies have used both PCA and DA to depict results of shape group designations, whether it is to compare a population to type or neotype specimens (e.g., [Stoermer & Ladewski 1982](#), [Stoermer et al. 1984](#), [Theriot & Ladewski 1986](#)) or find the degree of shape similarity in specimens from geographically separate places (e.g., [Stoermer et al. 1986](#)) or find the shape differences of specimens within a species complex (e.g., [Pappas & Stoermer 2003](#)).

Legendre polynomials, P_n , have been shown to be valuable in interpretation and matching numerical outcome to physical shape descriptions. Legendre polynomials and their coefficients represent sequential addition of valve shape features and that this representation could be interpreted within the context of principal components or canonical eigenvectors. Because of allometric considerations, and that all size influence cannot be removed (e.g., [Mosimann 1970](#)), size classes may implicitly exist within the total shape variation gradient, and this is expressed as size reduction series indicated on the first principal component (PC1) in ordinated shape space (e.g., [Stoermer & Ladewski 1982](#), [Stoermer et al. 1986](#), [Theriot & Ladewski 1986](#)).

From past studies, assignment of shape and symmetry descriptions to P_n with regard to PCA or DA may be compiled into a general scheme. While PC1 is associated with size reduction, the second principal component (PC2) has been found to define transapical asymmetry in *Gomphoneis* ([Stoermer & Ladewski 1982](#), [Stoermer et al. 1984](#)), *Didymosphenia* ([Stoermer et al. 1986](#)), and *Meridion* ([Rhode et al. 2001](#)). For *Eunotia* ([Steinman & Ladewski 1987](#)), PC2 defines apical asymmetry. Mathematically, this is the same property if an apically symmetric form is rotated 90°. Both heteropolar and dorsiventral isopolar valve shapes exhibit asymmetry ([Barber & Haworth 1981](#)). For *Tabellaria*, PC2 signifies the change in degree of constriction between the mid-valve area and the ends ([Theriot & Ladewski 1986](#)). This form is apically and transapically symmetric, so that PC2 also defines bilateral isopolar shapes ([Barber & Haworth 1981](#)). Overall, PC2 may be interpreted to be an initial gradient of shape symmetry.

For the third principal component (PC3), clavateness is more evident in *Gomphoneis* ([Stoermer & Ladewski 1982](#), [Stoermer et al. 1984](#)), *Didymosphenia* ([Stoermer et al. 1986](#)), and *Meridion* ([Rhode et al. 2001](#)), just as crescenticness is more pronounced for *Eunotia* ([Steinman & Ladewski 1987](#)). Increase in mid-valve area dorsally is associated with PC3, while change in shape of the ventral margin is associated with the fourth principal component (PC4) for *Eunotia* ([Steinman & Ladewski 1987](#)). Elongation of the valve dorsally and an increase in apex size and ventral inflation for *Eunotia* is characterized by the fifth, sixth and seventh principal components (PC5, PC6 and PC7, respectively) ([Steinman & Ladewski 1987](#)). In *Gomphoneis* ([Stoermer & Ladewski 1982](#), [Stoermer et al. 1984](#)), PC4-PC7 are associated with changes in headpole and undulations of the valve margin. As more and more PCs are considered, more and more aspects of valve shape and symmetry are accounted for.

The *Cymbella cistula* species complex also showed how physical morphological description could be matched to Legendre polynomials and interpreted as such in canonical shape space (Pappas & Stoermer 2003). As each subsequent P_n is considered, stepwise shape changes may be documented. For the P_0 and P_1 , semicircularity is first evident. This coincides with asymmetry associated with PC2. P_2 through P_4 provide further shape development dorsally as crescentic, arcuate and semilanceolate forms become more apparent, and this coincides with dorsal crescenticness associated with PC3. Expansion of the mid-valve area ventrally as an increase in gibbousness is produced in association with P_6 through P_8 . This coincides with ventral shape changes associated with PC4-PC7. *Cymbella* shape changes are correlated with the first two DA eigenvectors, with most of the correlation being attributable to the first canonical axis. The second canonical axis is correlated with P_8 , indicating that as one of the last polynomials used for curve-fitting, definition of the mid-valve area ventrally was produced to be used as a differentiating factor among specimens.

For *Fragilariforma* (Kingston & Pappas 2009), Legendre polynomial P_6 was used to begin to distinguish valve margins. With the addition of P_8 and P_9 , biundulate and triundulate margins were evident, and each inflation of biundulate form exhibited biundulation, or tetraundulation was present. The differences in valve margin and number of inflations were correlated to the first canonical eigenvector in DA. The second canonical eigenvector in DA was correlated with P_3 through P_5 and defined rostrateness of valve ends.

To summarize in the most general way, PC1 is associated with size reduction, PC2 is associated with valve asymmetry, PC3 is associated with clavateness or dorsally crescenticness, PC4 is associated with changes in valve shape ventrally, and along with PC5-PC7, changes in shape of valve ends and additional undulations on the valve margin. For Legendre polynomials, P_1 is associated with PC2, P_2 through P_5 are associated with PC3, and P_6 through P_8 are associated with PC4-PC7. Additional polynomials and PCs will provide more incremental changes in valve shape.

Potentially applicable combination and new morphometric methods

From all that has been described, it should be evident that all morphometric methods are concerned with using coordinates on the surface or outline of a diatom valve. It is not necessarily the case that a large difference exists between landmark-based and outline methods as it is in the difficulty in choosing points on a diatom valve surface (=“landmarks”) in order to represent that diatom valve in a morphologically (biologically) meaningful way with respect to additional analysis. Legendre shape analysis has been shown repeatedly to be a valuable morphometric tool. It would be desirable to expand its use in conjunction with other methods.

Recognition by Bookstein (1997) and others (e.g., MacLeod 1999, Sheets et al. 2004) that landmark-based and outline methods have more in common than they are different in morphometric analysis is a testament to the necessity of developing applications of combination methods to cover a broad spectrum of cases. Adaptations of landmark-based and outline combination methods include edgels (Bookstein & Green 1993, Little & Mardia 1996) and creases (Bookstein 2000), as well as extended eigenshape analysis (MacLeod 1999), finite element analysis (e.g., Singh et al. 1997, Lestrel 2000) and wavelet analysis (e.g., Takemura et al. 2004, Lestrel 2000). These are a few of the approaches that have been used recently in morphometrics. Although many of these methods have been applied in the study of human craniofacial structures, perhaps such methods could be altered or adapted to studies of organisms such as diatoms.

Orthogonal polynomials other than Legendre polynomials may be used in shape analysis. More care must be taken since weight functions are not equal to 1 when using Chebyshev, Zernike and other orthogonal polynomials. This methodology has yet to be worked out. It remains to be seen if other orthogonal polynomials can be matched to physical descriptions in a stepwise fashion or interpretable with respect to eigenvectors defining n -dimensional shape space.

Orthogonal moments functions, which involve the discrete form of Legendre, Chebyshev, or Zernike polynomials (e.g., [Novotni & Klein 2003](#)) as a basis, may provide a way to take advantage of concepts from both object recognition studies and shape analysis in devising a possible new method of morphometric analysis of diatoms. Using orthogonal moments has been found to provide a method with minimal noise in image reconstruction (e.g., [Mukundan et al. 2001](#)) and may be modified as orthonormal moments for use as shape features ([Mukundan 2004](#)). A speculative method such as this would need much work from the ground up and rigorous testing to apply to studies of diatoms.

Another method worth mentioning involves shape modeling (e.g., [Hart 1998](#), [Biasotti et al. 2003](#)) based on concepts from algebraic topology, including Morse theory (e.g., [Milnor 1963](#)), homotopy (= deformability) (e.g., [Kahn 1995](#)) and homology (= equivalence relation, in a topological sense) (e.g., [Kahn 1995](#)). This kind of modeling is graph-based rather than being based on geometric concepts. One example of a recent graph-based modeling study entails skeleton-based shape matching (e.g., [Sundar et al. 2003](#), [Zhu 2007](#)) using Reeb graphs ([Reeb 1946](#)). In fact, graph-matching has been used in diatom research ([Pappas 2011](#)). Much of the object recognition literature, including shape and pattern recognition, as evidenced by ADIAC, may prove to contain many useful concepts in developing new morphometric methods, provided that biologically meaningful information is obtained from such methods rather than merely topologically or other kinds of mathematically interesting results.

When talking about shape analysis using outline methods, it is implied that the entire outline of diatom valves are modeled. In relatively few studies, partial diatom outlines have been used to differentiate shapes and assign species group designations (e.g., [Pappas & Stoermer 2003](#), [Potapova & Hamilton 2007](#)). Partial shape extraction using contour segments based on the Fourier transform has been used as a shape matching technique for some time (e.g., [Gorman et al. 1988](#)). It may be a fruitful avenue of discovery if more studies using partial diatom valve outlines were undertaken, since an implication of such analysis would be to find the critical, and perhaps minimal, part of diatom valve shape that defines a species.

Other multivariate methods may be useful in ordinating shape coefficients. Non-linear PCA (such as correspondence analysis (e.g., [Hill 1974](#)) and non-linear DA (e.g., [Gifi 1990](#)) may provide a way to analyze data when the whole continuum of size classes is not present, as was the case for the Legendre shape analysis of *Tabellaria* (e.g., [Theriot & Ladewski 1986](#)). In addition, more use could be made of canonical variates analysis and MANOVA, and perhaps, other multivariate methods could be adapted for use in analyzing shape coefficients. In time, testing of adapted methods may fulfill the possibilities of devising new multivariate analytical techniques for use in shape analysis.

Actual application of new or additional modified morphometric methods is even more likely when more is known about the development, life cycle, or ontogeny of diatoms. It is difficult to generalize a method to a group of organisms, such as diatoms, when very little is known about something so fundamentally important, e.g., as reproduction, is not considered as a means to judge the effectiveness of a morphometric method. Moreover, for a more useful application of morphometric methods, a more clear understanding is required of changes in shape and morphological features of diatom valves from auxospore to initial cell to vegetative cell, exactly how much stepwise shape change occurs in each vegetative cell division, and how this is related to and identifiable at the generic, and preferably, the species level.

No one method will work for all times and all places, and no one method is superior to another. Although there are problems with using landmark-based methods and outline shape analysis, it is possible that these problems may be overcome when more is known about the development, functionality, and phylogeny of diatom groups. Alternatively, as older methods are modified or current methods are combined or brand-new methods are created, new ways of using quantitative morphometrics will surely become available to diatom researchers for use in their studies. Moreover, we have only reviewed 2D methods here. With advances in technology and computing power, 3D applications are in the offing, which is an entirely separate topic in its own

right. Morphometric methods have been a valuable tool in diatom taxonomic and classification studies and will continue to be increasingly important as ways to extend our understanding of these important microorganisms.

Acknowledgements

We dedicate this paper to Nina Strelnikova for her life-long contributions to diatom research. We also thank Ed Theriot for reading and commenting on a previous version of this manuscript.

References

- Abramowitz, M. & Stegun, I.A. (eds) (1972): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, U.S. Department of Commerce, National Bureau of Standards, Applied Mathematics Series 55 – Washington, D.C. 1060 pp.
- Adams, D.C. & Rosenberg, M.S. (1998): Partial warps, phylogeny, and ontogeny: a comment on Fink and Zelditch (1995). – *Syst. Biol.* **47**: 168–173.
- Adams, D.C., Rohlf, F.J. & Slice, D.E. (2004): Geometric morphometrics: ten years of progress following the ‘revolution’. – *Italian J. Zool.* **71**: 5–16.
- Almeida, S.P. & Eu, J.K.T. (1976): Water pollution monitoring using matched spatial filters. – *Applied Optics* **15**: 510–515.
- Barber, H.G. & Haworth, E.Y. (1981): *A Guide to the Morphology of the Diatom Frustule*. – Freshwat. Biol. Assoc. Sci. Publ. **44**: 112 pp. The Ferry House, Ambleside, Cumbria, UK.
- Bennett, J.R. & MacDonald, J.S. (1975): On the measurement of curvature in a quantized environment. – *IEEE Transactions on Computers* **24**: 803–820.
- Beszteri, B., Ács, É. & Medlin, L. (2005): Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyta). – *Eur. J. Phyc.* **40**: 89–103.
- Biasotti, S., Marini, S., Mortara, M., Patanè, G., Spagnuolo, M. & Falcidieno, B. (2003): 3D shape matching through topological structures. – In: Nyström et al. (eds): *Lecture Notes in Computer Science*, Vol. **2886**: 194–203. Springer Verlag, Berlin.
- Bookstein, F.L. (1978): *The Measurement of Biological Shape and Shape Change*. – *Lecture Notes in Biomathematics* **24**: 191 pp. Springer Verlag, New York.
- Bookstein, F.L. (1986): Size and shape spaces for landmark data in two dimensions (with discussion). – *Statistical Sci.* **1**: 181–242.
- Bookstein, F.L. (1989): Principal warps: thin plate splines and the decomposition of deformations. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**: 567–585.
- Bookstein, F.L. (1991): *Morphometric Tools for Landmark Data*. – Cambridge Univ. Press. 435 pp.
- Bookstein, F.L. (1996): *Landmark methods for forms without landmarks: localizing group differences in outline shape*. – In: Amini, A., Bookstein, F.L. & Wilson, D. (eds): *Proceedings of the workshop on mathematical methods in biomedical image analysis*, San Francisco, pp. 279–289. IEEE Computer Society.
- Bookstein, F.L. (1997): *Landmark methods for forms without landmarks: morphometrics of group differences in outline shape*. – *Medical Image Analysis* **1**: 225–243.
- Bookstein, F.L. (2000): *Creases as local features of deformation grids*. – *Medical Image Analysis* **4**: 93–110.
- Bookstein, F.L., Chernoff, B., Elder, R., Humphries, J., Smith, G. & Strauss, R. (1985): *Morphometrics in Evolutionary Biology*. – Acad. Nat. Sci. Philadelphia, Spec. Publ. **15**: 277 pp.
- Bookstein, F.L. & Green, W.D.K. (1993): A feature space for edgels in images with landmarks. – *J. Math. Imaging and Vision* **3**: 231–261.
- Bookstein, F.L., Streissguth, A.P., Sampson, P.D., Connor, P.D. & Barr, H.M. (2002): Corpus callosum shape and neuropsychological deficits in adult males with heavy fetal alcohol exposure. – *Neuroimage* **15**: 233–251.
- Brown, J.W. & Churchill, R.V. (2001): *Fourier Series and Boundary Value Problems*. 6th ed. – McGraw-Hill, New York. 344 pp.
- Burke, J.F. (1970): *A review of the genus Aulacodiscus*. – Staten Island Institute of Arts and Sciences, 306, 314. New York.
- Cairns, Jr., J., Dickson, K.L., Slocumb, J.P., Almeida, S.P., Eu, J.K.T., Liu, C.Y.C. & Smith, H.F. (1974): *Microcosm pollution monitoring*. – In: Hemphill, D.D. (ed.): *Trace Substances in Environmental Health –VIII*: 223–228. University of Missouri, Columbia.

- Cairns, Jr., J., Dickson, K.L., Slocumb, J.P., Almeida, S.P. & Eu, J.K.T. (1977): Biological monitoring of aquatic community structure using a computer interfaced laser system. – In: Alabaster, J.S. (ed.): *Biological Monitoring of Inland Fisheries*: 143–150. Appl. Sci. Publ., Ltd., Essex, England.
- Cairns, JR. J., Almeida, S.P. & Fujii, H. (1982): Automated identification of diatoms. – *Bioscience* **32**: 98–102.
- Canny, J. (1986): A computational approach to edge detection. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8**: 679–698.
- Cartmill, M. (1994): A critique of homology as a morphological concept. – *Am. J. Phys. Anthropol.* **94**: 115–124.
- Ciobanu, A. & du Buf, H. (2002): Identification by contour profiling and Legendre polynomials. – In: du Buf, H. & Bayer, M.M. (eds): *Automatic Diatom Identification and Classification*: 167–185. World Sci., New Jersey.
- Cooley, W.W. & Lohnes, P.R. (1971): *Multivariate Data Analysis*. – John Wiley & Sons, New York. 364 pp.
- Cracraft, J. (1967): Comments on homology and analogy. – *Syst. Zool.* **16**: 355–359.
- Dryden, I.L. & Mardia, K.V. (1991): General shape distributions in a plane. – *Adv. Appl. Probab.* **23**: 259–276.
- Dryden, I.L. & Mardia, K.V. (1998): *Statistical Shape Analysis*. – J. Wiley & Sons, Chichester, UK. 347 pp.
- Edwards, R.E. (1967): *Fourier Series: A Modern Introduction*, Vol. 1. – Holt, Rinehart and Winston, Inc., New York. 211 pp.
- Ehrlich, R., Pharr, R.B., JR. & Healy-Williams, N. (1983): Comments on the validity of Fourier descriptors in systematics: a reply to Bookstein et al. – *Syst. Zool.* **32**: 202–206.
- Falasco, E., Blanco, S., Bona, F., Gomá, J., Hlúbíková, D., Novais, M.H., Hoffmann, L. & Ector, L. (2009): Taxonomy, morphology and distribution of the *Sellaphora stroemii* complex (Bacillariophyceae). – *Fottea* **9**: 243–256.
- Ferson, S., Rohlf, F.J. & Koehn, R.K. (1985): Measuring shape variation of two-dimensional outlines. – *Syst. Zool.* **34**: 59–68.
- Fischer, S. & Bunke, H. (2002): Identification using classical and new features in combination with decision tree ensembles. – In: du Buf, H. & Bayer, M.M. (eds): *Automatic Diatom Identification and Classification*: 109–140. World Sci., New Jersey.
- Fischer, S., Shahbazkia, J.R. & Bunke, H. (2002): Contour extraction. – In: du Buf, H. & Bayer, M.M. (eds): *Automatic Diatom Identification and Classification*: 93–107. World Sci., New Jersey.
- Forero-Vargas, M., Redondo, R. & Cristobal, G. (2003): Diatom screening and classification by shape analysis. – In: García, N., Martínez, J.M. & Salgado, L. (eds): *Lecture Notes in Computer Science*. Vol. **2849**: 58–65. Springer Verlag, Berlin.
- Geissler, U. (1970a): Die Variabilität der Schalenmerkmale bei der Diatomeen. – *Nova Hedwigia* **19**: 623–773.
- Geissler, U. (1970b): Die Schalenmerkmale der Diatomeen. Ursachen ihrer Variabilität und Bedeutung für die Taxonomie. – *Nova Hedwigia*, Beih. **31**: 511–535.
- Geissler, U. (1982): Experimentelle Untersuchungen zur Variabilität der Schalenmerkmale bei einigen zentrischen Süßwasser-Diatomeen. I. Der Einfluss unterschiedlicher Salzkonzentrationen auf den Valva-Durchmesser von *Stephanodiscus hantzschii* Grunow. – *Nova Hedwigia*, Beih. **73**: 211–247.
- Genkal, S.I. & Kuzmin, G.V. (1979): Biometric analysis of main structural elements of valves in the species of the genus *Stephanodiscus* Ehr. (Bacillariophyta). – *Bot. Zhurnal* **64**: 1237–1244.
- Gevirtz, J.L. (1976): Fourier analysis of bivalve outlines: implications on evolution and autecology. – *Math. Geol.* **8**: 151–163.
- Ghiselin, M.T. (1969): The distinction between similarity and homology. – *Syst. Zool.* **18**: 148–149.
- Gifi, A. (1990): *Nonlinear Multivariate Analysis*. – Chichester, New York. 579 pp.
- Goldman, N., Paddock, T.B.B. & Shaw, K.M. (1990): Quantitative analysis of shape variation in populations of *Surirella fastuosa*. – *Diatom Res.* **5**: 25–42.
- Goodall, C.R. (1991). Procrustes methods in the statistical analysis of shape. – *J. Roy. Stat. Soc. Ser. B* **53**: 285–339.
- Goodall, C.R. & Mardia, K.V. (1993): Multivariate aspects of shape theory. – *Ann. Statistics* **21**: 848–866.
- Gorman, J.W., Mitchell, O.R. & Kuhl, F.P. (1988): Partial shape recognition using dynamic programming. – *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**: 257–266.
- Gower, J.C. (1975). Generalized procrustes analysis. – *Psychometrika* **40**: 33–51.
- Green, P.E. & Carroll, J.D. (1978): *Analyzing Multivariate Data*. – The Dryden Press, Hinsdale, Illinois. 519 pp.

- Green, W.D.K. (1995): Spline-based deformable models. – In: Melter, R.A., Wu, A.Y., Bookstein, F.L. & Green, W.D.K. (eds): *Vision Geometry IV, SPIE Proceedings*, Vol. **2573**: 290–301. The Internat. Society for Optical Engineering (Society of Photo-Optical Instrumentation Engineers), San Diego, California.
- Hart, J.C. (1998): Morse theory for implicit surface modeling. – In: Hege, H.-C. & Polthier, K. (eds): *Mathematical Visualization: 257–268*. Springer Verlag, Berlin.
- Hicks, Y.A., Marshall, D., Rosin, P.L., Martin, R.R., Mann, D.G. & Droop, S.J.M. (2006): A model of diatom shape and texture for analysis, synthesis and identification. – *Machine Vision Appl.* **17**: 297–307.
- Hill, M.O. (1974): Correspondence analysis: a neglected multivariate method. – *Appl. Stat.* **23**: 340–354.
- Holmes, R.W. & Reimann, B.E.F. (1966): Variation in valve morphology during the life cycle of the marine diatom *Coscinodiscus concinnus*. – *Phycologia* **5**: 233–244.
- Jalba, A.C., Wilkinson, M.H.F., Roerdink, J.B.T.M., Bayer, M.M. & Juggins, S. (2005): Automatic diatom identification using contour analysis by morphological curvature scale spaces. – *Machine Vision Appl.* **16**: 217–228.
- Johansen, J.R. & Theriot, E.C. (1987): The relationship between valve diameter and number of central fuloportulae in *Thalassiosira weissflogii* (Bacillariophyceae). – *J. Phyc.* **23**: 663–665.
- Johnston, E.M. & Stoermer, E.F. (1976): Computer analysis of phytoplankton cell images. – *The Microscope* **24**: 181–187.
- Jongman, R.H., Ter Braak, C.J.F. & van Tongeren, O.F.R. (1995): *Data Analysis in Community and Landscape Ecology*. – Cambridge Univ. Press, New York. 299 pp.
- Kahn, D.W. (1995): *Topology: An Introduction to the Point-Set and Algebraic Areas*. – Dover Publ., Inc., New York. 217 pp.
- Kendall, D.G. (1977): The diffusion of shape. – *Adv. Appl. Prob.* **9**: 428–430.
- Kendall, D.G. (1984): Shape manifolds, procrustean metrics, and complex projective spaces. – *Bull. London Math. Soc.* **16**: 81–121.
- Kendall, D.G., Barden, D., Carne, T.K. & Le, H. (1999): *Shape and Shape Theory*. – John Wiley & Sons, Ltd., Chichester, UK. 306 pp.
- Kincaid, D.T. & Schneider, R.B. (1983): Quantification of leaf shape with a microcomputer and Fourier transform. – *Can. J. Bot.* **61**: 2333–2342.
- Kingston, J.C. & Pappas, J.L. (2009): Quantitative shape analysis as a diagnostic and prescriptive tool in determining *Fragilariforma* (Bacillariophyta) taxon status. – *Nova Hedwigia, Beih.* **135**: 103–119.
- Kobayashi, A., Tanaka, J. & Nagumo, T. (1998): Morphological and taxonomical study of *Arachnoidiscus ornatus* Ehb. (Bacillariophyceae) in Japan. – *Diatom* **14**: 25–34.
- Kociolek, J.P. & Stoermer, E.F. (1988): Taxonomy, ultrastructure, and distribution of *Gomphoneis herculeana*, *G. eriense* and closely related species. – *Proc. Acad. Nat. Sci. Philadelphia* **140**: 24–97.
- Kuhl, F.P. & Giardina, C.R. (1982): Elliptic Fourier features of a closed contour. – *Computer Graphics Image Processing* **9**: 3236–3258.
- Le, H. & Kume, A. (2000): The Fréchet mean shape and shape of means. – *Adv. Appl. Prob.* **32**: 101–113.
- Lestrel, P.E. (ed.) (1997): *Fourier Descriptors and their Applications in Biology*. – Cambridge Univ. Press, Cambridge, UK. 466 pp.
- Lestrel, P.E. (2000): Morphometrics for the Life Sciences. – In: Oxnard, C.E. (ed.): *Recent Advances in Human Biology*, Vol. **7**: 1–261. World Sci., Singapore.
- Little, J. & Mardia, K. (1996): Edgels and tangent planes in image warping. – In: Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P. & Slice, D.E. (eds): *Advances in Morphometrics: 263–270*. Plenum Press, New York.
- Lohmann, G.P. (1983): Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. – *Math. Geol.* **15**: 659–672.
- Loke, R.E., Bayer, M.M., Mann, D.G. & du Buf, J.M.H. (2002): Diatom recognition by convex and concave contour curvature. – *Oceans '02/IEEE* **4**: 2457–2465.
- Loke, R.E. & du Buf, H. (2002): Identification by curvature of convex and concave segments. – In: du Buf, H. & Bayer, M.M. (eds.): *Automatic Diatom Identification and Classification: 141–165*. World Sci., New Jersey.
- MacLeod, N. (1999): Generalizing and extending the eigenshape method of shape space visualization and analysis. – *Paleobiology* **25**: 107–138.
- MacLeod, N., O'Neill, M. & Walsh, S.A. (2007): A comparison between morphometric and artificial neural network approaches to the automated species recognition problem in systematics. – In: Curry, G.B. &

- Humphries, C.J. (eds): Biodiversity Databases: Techniques, Politics, and Applications: 37–62. CRC Press, Boca Raton, Florida.
- Mahalanobis, P.C. (1948): Historical note on the D^3 -statistic. – *Sankhya*, **9**: 237.
- Manly, B.F.J. (2005): *Multivariate Statistical Methods: A Primer*, 3rd edition. – Chapman and Hall, London. 214 pp.
- Mann, D.G., McDonald, S.M., Bayer, M.M., Droop, S.J.M., Chepurnov, V.A., Loke, R.E., Ciobanu, R.E. & du Buf, J.M.H. (2004): The *Sellaphora pupula* species complex (Bacillariophyceae): morphometric analysis, ultrastructure and mating data provide evidence for five new species. – *Phycologia* **43**: 459–482.
- Marcus, L.F., Hingst-Zaher, E. & Zaher, H. (2000): Application of landmark morphometrics to skulls representing the orders of living mammals. – *Hystrix* **11**: 27–47.
- Mardia, K.V. & Dryden, I.L. (1989a): Shape distributions for landmark data. – *Adv. Appl. Prob.* **21**: 742–755.
- Mardia, K.V. & Dryden, I.L. (1989b): The statistical analysis of shape data. – *Biometrika* **76**: 271–281.
- Mardia, K.V. & Dryden, I.L. (1994): Shape averages and their bias. – *Adv. Appl. Prob.* **26**: 334–340.
- Mardia, K.V. & Patrangenaru, V. (2005): Directions and projective spaces. – *Ann. Statistics* **33**: 1666–1669.
- Mardia, K.V., Bookstein, F.L. & Moreton, I.J. (2000): Statistical assessment of bilateral symmetry of shapes. – *Biometrika* **87**: 285–300.
- Milnor, J. (1963): *Morse Theory*. – Annals of Mathematical Studies, Princeton Univ. Press, Princeton, New Jersey. 153 pp.
- Mizuno, M. (1987): Morphological variation of the attached diatom *Cocconeis scutellum* var. *scutellum* (Bacillariophyceae). – *J. Phyc.* **23**: 591–597.
- Mosimann, J.E. (1970): Size allometry: size and shape variables with characterization of the lognormal and generalized gamma distributions. – *J. Am. Statistical Assoc.* **65**: 930–948.
- Mou, D. & Stoermer, E.F. (1992): Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors. – *J. Phyc.* **28**: 386–395.
- Mukundan, R. (2004): Some computational aspects of discrete orthonormal moments. – *IEEE Transactions on Image Processing* **13**: 1055–1059.
- Mukundan, R., Ong, S.H. & Lee, P.A. (2001): Discrete vs. continuous orthogonal moments for image analysis. – In: Arabnia, H.R. & Mun, Y. (eds): *Proceedings of the International Conference on Imaging Science, System and Technology*, Vol. 1: 23–29. CSREA Press, Las Vegas, Nevada.
- Novais, M.H., Blanco, S., Hlúbíková, D., Falasco, E., Gomá, J., Delgado, C., Ivanov, P., Ács, É., Morais, M., Hoffmann, L. & Ector, L. (2009): Morphological examination and biogeography of the *Gomphonema resenstockianum* and *G. teregestinum* species complex (Bacillariophyceae). – *Fottea* **9**: 257–274.
- Novotni, M. & Klein, R. (2003): 3D Zernike descriptors for content based shape retrieval. – In: Elber, G. & Shapir, V. (eds): *Proceedings of the Eighth ACM Symposium on Solid Modeling and Applications*: 216–225. ACM Press, New York.
- Noy-Meir, I. (1973): Data transformations in ecological ordination: I. Some advantages of non-centering. – *J. Ecol.* **61**: 329–341.
- Noy-Meir, I., Walker, D. & Williams, W.T. (1975): Data transformations in ecological ordination. II. On the meaning of data standardization. – *J. Ecol.* **63**: 779–800.
- Pappas, J.L. (2000): Fourier shape analysis and shape group determination by principal component analysis and fuzzy measure theory of *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae) from the Great Lakes. – Doctoral Diss., Univ. Michigan, Ann Arbor, Michigan. 170 pp.
- Pappas, J.L. (2006): Biological taxonomic problem solving using fuzzy decision-making analytical tools. – *Fuzzy Sets and Systems* **157**: 1687–1703.
- Pappas, J.L. (2011): Graph matching a skeletonized theoretical morphospace with a cladogram for gomphonemoid-cymbelloid diatoms (Bacillariophyta). – *J. Biol. Syst.* **19**: 47–70.
- Pappas, J. L. & Stoermer, E.F. (2001): Fourier shape analysis and fuzzy measure shape group differentiation of Great Lakes *Asterionella* Hassall (Heterokontophyta, Bacillariophyceae). – In: Economou-Amilli, A. (ed.): *Proceedings of the Sixteenth International Diatom Symposium*: 485–501. Amvrosiou Press, Univ. Athens, Greece.
- Pappas, J. L. & Stoermer, E.F. (2003): Legendre shape descriptors and shape group determination of specimens in the *Cymbella cistula* species complex. – *Phycologia* **42**: 90–97.
- Pappas, J.L., Fowler, G.W. & Stoermer, E.F. (2001): Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae). – *Phycologia* **40**: 440–456.
- Perez, S.I., Bernal, V. & Gonzalez, P.N. (2006): Differences between sliding semi-landmark methods in

- geometric morphometrics, with an application to craniofacial and dental variation. – *J. Anatomy* **208**: 769–784.
- Persoon, E. & Fu, K.S. (1977): Shape discrimination using Fourier descriptors. – *IEEE Transactions on Systems, Man, and Cybernetics* **7**: 170–179.
- Potapova, M. & Hamilton, P.B. (2007): Morphological and ecological variation within the *Achnanthydium minutissimum* (Bacillariophyceae) species complex. – *J. Phycol.* **43**: 561–575.
- Reeb, G. (1946): Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. – *C. R. Acad. Sci. Paris* **222**: 847–849.
- Reimer, C.W. (1954): Re-evaluation of the diatom species *Nitzschia frustulum* (Kütz.) Grun. – *Butler Univ. Bot. Stud.* **11**: 178–191.
- Rhode, K.M., Pappas, J.L. & Stoermer, E.F. (2001): Quantitative analysis of shape variation in type and modern populations of *Meridion* (Bacillariophyceae). – *J. Phycol.* **37**: 176–183.
- Rivera, P.S & Barrales, H.L. (1994): *Asteromphalus sarcophagus* Wallich and other species of the genus off the coast of Chile. – *Mem. Calif. Acad. Sci.* **17**: 37–54.
- Rohlf, F.J. (1990): Fitting curves to outlines. – In: Rohlf, F.J. & Bookstein, F.L. (eds): *Proceedings of the Michigan Morphometrics Workshop, Spec. Publ.* **2**: 167–177. Univ. Michigan Mus. Zool., Ann Arbor, Michigan.
- Rohlf, F.J. (1996): Morphometric spaces, shape components and the effects of linear transformations. – In: Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P. & Slice, D.E. (eds): *Advances in Morphometrics*: 117–129. Plenum Press, New York.
- Rohlf, F.J. (1998): On applications of geometric morphometrics to studies of ontogeny and phylogeny. – *Syst. Biol.* **47**: 147–158.
- Rohlf, F.J. & Archie, J.W. (1984): A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera: Culicidae). – *Syst. Zool.* **33**: 302–317.
- Rohlf, F.J. & Bookstein, F.L. (eds.) (1990): *Proceedings of the Michigan Morphometrics Workshop, Spec. Publ.* **2**. – Univ. Michigan Mus. Zool., Ann Arbor, Michigan. 380 pp.
- Rohlf, F.J. & Slice, D. (1990): Extensions of the Procrustes method for the optimal superimposition of landmarks. – *Syst. Zool.* **39**: 40–59.
- Sahoo, P.K., Soltani, S. & Wong, K.C. (1988): A survey of thresholding techniques. – *Computer Vision and Graphics Image Processing* **41**: 279–295.
- Santos, L.M. & du Buf, H. (2002): Identification by Gabor features. – In: du Buf, H. & Bayer, M.M. (eds): *Automatic Diatom Identification and Classification*: 187–220. World Sci., New Jersey.
- Sheets, H.D., Keonho, K. & Mitchell, C.E. (2004): A combined landmark and outline-based approach to ontogenetic shape change in the Ordovician trilobite *Triarthrus becki*. – In: Elewa, A. (ed.): *Applications of Morphometrics in Paleontology and Biology*: 67–81. Springer Verlag, New York.
- Sheets, H.D., Covino, K.M., Panasiewicz, J.M. & Morris, S.R. (2006): Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. – *Frontiers in Zoology* **3**: 15–26.
- Singh, G.D., McNamara, J.A., JR. & Lozanoff, S. (1997): Finite element analysis of the cranial base in subjects with class III malocclusion. – *British J. Orthodontics* **24**: 103–112.
- Slice, D.E. (2001): Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space. – *Syst. Biol.* **50**: 141–149.
- Slice, D.E., Bookstein, F.L. Marcus, L.F. & Rohlf, F.J. 1996. A Glossary for Geometric Morphometrics. – In: Marcus, L.F., Corti, M., Loy, A., Naylor, G.J.P. & Slice, D.E. (eds): *Advances in Morphometrics*: 531–551. Plenum Press, New York. (Online at: <http://dipbau.bio.uniroma1.it/web/Docenti/Docente373/Morfometria-geometrica/glossary-of-geometric-morphometrics.pdf> for a download, or type in “morphometrics glossary” in a search engine box to get a webpage version of the glossary).
- Small, C.G. (1988): Techniques of shape analysis on sets of points. – *Int. Statistical Rev.* **56**: 243–257.
- Small, C.G. (1996): *The Statistical Theory of Shape*. – Springer Verlag, New York. 227 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.
- Steinman, A.D. & Sheath, R.G. (1984): Morphological variability of *Eunotia pectinalis* (Bacillariophyceae) in a softwater Rhode Island stream and in culture. – *J. Phycol.* **20**: 266–276.
- Stoermer, E.F. & N.A. ANDRESEN (2006): Atypical *Tabularia* in coastal Lake Erie, USA. In: N. Ognjanova-Rumernova & K. Manoylov (eds.): *Advances in Phycological Studies, Festschrift in honour of Prof. Dobrina Temniskova-Topalova*: 353-363. Pensoft Publishers & University Publishing House, Sofia-Moscow.

- Stoermer, E.F. & Ladewski, T.B. (1982): Quantitative analysis of shape variation in type and modern populations of *Gomphonopsis herculeana*. – *Nova Hedwigia*, Beih. **73**: 347–386.
- Stoermer, E.F., Ladewski, T.B. & Kociolek, J.P. (1984): Further observations on *Gomphonopsis*. – In: Richard, M. (ed.): *Proceedings of the Eighth International Diatom Symposium*: 205–213. Koeltz Scientific, 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.
- Szőkefalvi-Nagy, B. (1965): *Introduction to Real Functions and Orthogonal Expansions*. – Oxford Univ. Press, New York. 447 pp.
- Takemura, C.M., Cesar, R.M., Jr., Arantes, R.A.T., Da, L., Costa, F., Hingst-Zaher, E., Bonato, V. & dos Reis, S.F. (2004): Morphometrical data analysis using wavelets. – *Real-Time Imaging* **10**: 239–250.
- Theriot, E. (1987): Principal component analysis and taxonomic interpretation of environmentally related variation in silicification in *Stephanodiscus* (Bacillariophyceae). – *British Phycol. J.* **22**: 359–373.
- Theriot, E. (1992): Clusters, species concepts, and morphological evolution of diatoms. – *Syst. Biol.* **41**: 141–157.
- Theriot, E. & Ladewski, T.B. (1986): Morphometric analysis of shape of specimens from the neotype of *Tabellaria flocculosa* (Bacillariophyceae). – *Am. J. Bot.* **73**: 224–229.
- Theriot, E.C. & Stoermer, E.F. (1982): Principal component analysis and character variation in *Stephanodiscus niagarae* Ehrenb.: morphological variation relative to lake trophic status. – In: Mann, D.G. (ed.): *Proceedings of the 6th International Diatom Symposium*: 97–111. O. Koeltz, Koenigstein.
- Theriot, E. & Stoermer, E.F. (1984): Principal component analysis of *Stephanodiscus*: observations on two new species from the *Stephanodiscus niagarae* complex. – *Bacillaria* **7**: 37–58.
- Theriot, E.C., Håkansson, H. & Stoermer, E.F. (1988): Morphometric analysis of *Stephanodiscus alpinus* (Bacillariophyceae) and its morphology as an indicator of lake trophic status. – *Phycologia* **27**: 485–493.
- Thompson, D'A.W. (1917): *On Growth and Form*. – Cambridge Univ. Press, Cambridge, UK. 793 pp.
- Thompson, D'A.W. (1942): *On Growth and Form*, 2nd edition. – Cambridge Univ. Press, Cambridge, UK. 1116 pp.
- Tropper, C.B. (1975): Morphological variation of *Achnanthes hauckiana* (Bacillariophyceae) in the field. – *J. Phycol.* **11**: 297–302.
- Veselá, J., Neustupa, J., Pichrtová, M. & Poulíčková, A. (2009): Morphometric study of *Navicula* morphospecies (Bacillariophyta) with respect to diatom life cycle. – *Fottea* **9**: 307–316.
- Weinberger, H.F. (1995): *A First Course in Partial Differential Equations with Complex Variables and Transform Methods*. – Dover Publications, Inc., New York. 446 pp.
- Wendker, S. (1990): Morphologische Untersuchungen an Populationen aus dem Formenkreis um *Nitzschia frustulum* (Kützing) Grunow. – *Diatom Res.* **5**: 179–187.
- Wilkinson, M.H.F., Jalba, A.C., Urbach, E.R. & Roerdink, J.B.T.M. (2002): Identification by mathematical morphology. – In: du Buf, H. & Bayer, M.M. (eds): *Automatic Diatom Identification and Classification*: 221–244. World Sci., Singapore.
- Wood, A.M., Lande, R. & Fryxell, G.A. (1987): Quantitative genetic analysis of morphological variation in an Antarctic diatom grown at two light intensities. – *J. Phycol.* **23**: 42–54.
- Yunker, J.L. & Ehrlich, R. (1977): Fourier biometrics: harmonic amplitudes as multivariate shape descriptors. – *Syst. Zool.* **26**: 336–342.
- Zahn, C.T. & Roskies, R.Z. (1972): Fourier descriptors for plane closed curves. – *IEEE Transactions on Computers* **21**: 269–281.
- Zelditch, M.L., Swiderski, D.L., Sheets, D.H. & Fink, W.L. (2004): *Geometric Morphometrics for Biologists: A Primer*. – Elsevier Academic Press, London. 452 pp.
- Zhu, X. (2007): Shape recognition based on skeleton and support vector machines. – In: Huang, D.-S., Heutte, L. & Loog, M. (eds): *Proceedings of the Third International Conference on Intelligent Computing, Communications in Computer and Information Science* **2**: 1035–1043. Springer Verlag, Berlin.
- Ziezold, H. (1977): On expected figures and a strong law of large numbers for random elements in quasi-metric spaces. – In: Czechoslovak Academy of Sciences, Prague (ed.): *Transaction of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Vol. A: 591–602. Reidel, Dordrecht.
- Ziezold, H. (1994): Mean figures and mean shapes applied to biological figure and shape distributions in the plane. – *Biometrical J.* **36**: 491–510.