Ashraf M.T. Elewa

Morphometrics for Nonmorphometricians

Springer

Lecture Notes in Earth Sciences   124

Ashraf M.T. Elewa
Editor

Morphometrics for Nonmorphometricians

Springer

Editor
Prof. Dr. Ashraf M.T. Elewa
Minia University
Fac. Science
Dept. Geology
Minia 61519
Egypt
aelewa@link.net
ashrafelewa@ymail.com

Foreword

The aim of this new edited volume is to present the various aspects of morphometrics in a way that is accessible to readers who might not be acquainted with the voluminous literature on this topic. Morphometrics is the quantitative study of organismic form. It attempts to quantify, precisely and as completely as practical, the information on size and shape that often is readily apparent to the investigator yet difficult to adequately characterize in numerical terms. Morphometric studies thus provide an initial step in the understanding of patterns of variation among individuals and groups of organisms, and typically form an important groundwork for analyses of structure, function, and evolution. As such, morphometrics has become central to the biological sciences.

The field of morphometrics has transitioned relatively smoothly through several different phases, from D'Arcy Thompson's (1917) extraordinary and influential treatise on growth and form, through the influx of algebraic and statistical methods related to eigenanalysis, cluster analysis, and multidimensional scaling, to direct landmark-based methods that echo Thompson's original intents and insights. The history of these ideas is reviewed by Richard Reyment, who has himself been an important contributor to and synthesizer of morphometric theory and practice.

Included among the other chapters are examples of applications of morphometric methods in palaeontology and neontology, fishery science, archaeology, and evolutionary ecology. The volume also includes reviews and perspectives on a number of currently important methodological topics in morphometrics, including approaches to the analysis of morphological variation and clinal variation, the quantification of curvature, discrimination and classification, automated typology, use of morphometric data in evolutionary and phylogenetic analysis, and the integration of traditional distance-based methods with geometric morphometrics.

The geometric morphometric methods that have been developed over the past several decades to extend beyond the limitations of traditional distance-based methods have become transformed into the new standard research protocol. As the technologies of measurement, analysis and display continue to improve; our current methods will continue to evolve. For those who are entering the field, it is hoped that

the papers in this volume may serve as a current set of landmarks in time, reflecting over past work and pointing toward the future.

Lubbock, Texas   Richard E. Strauss

Contents

Contributors

Ahmed A. Abdelhady Geology Department, Faculty of Science, Minia University, Box 61519, Minia, Egypt, alhady2003@yahoo.com

Juan A. Barceló Departament de Prehistòria, Facultat de Lletres, Universitat Autònoma de Barcelona, Campus Bellaterra, 08193 Cerdanyola, Barcelona, España juanantonio.barcelo@uab.es

Casper J. Breuker Evolutionary Developmental Biology Research Group, Sinclair Building, School of Life Sciences, Oxford Brookes University, Gipsy Lane, Headington, Oxford OX3 0BP, UK; Group of Evolutionary Biology, Department of Biology, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium, cbreuker@brookes.ac.uk

Esther Bühler Department of Biology, Institute of Ecology & Evolution, University of Bern, CH3012, Bern, Switzerland, e.buehler@pobox.ch

Steven X. Cadrin NOAA/UMass Cooperative Marine Education & Research Program, School for Marine Science, Science and Technology, 200 Mill Road, Suite 325, Fairhaven, MA 02719, USA, Steven.Cadrin@noaa.gov

Marcelo Cardillo IMHICIHU-CONICET. Saavedra 5 th floor, Buenos Aires, Argentina, marcelo.cardillo@gmail.com

Andrea Cardini Dipartimento di Biologia Animale, Universitá di Modena e Reggio Emilia, via Campi 213, 41100 Modena, Italy, alcardini@interfree.it, cardini@unimo.it

P. David Polly Department of Geological Sciences, Indiana University, Bloomington, IN 47405 USA, pdpolly@indiana.edu

Ashraf M.T. Elewa Geology Department, Faculty of Science, Minia University, Box 61519, Minia, Egypt, aelewa@link.net; ashrafelewa@ymail.com

Sarah Elton Hull York Medical School, The University of Hull, Hull HU6 7RX, UK, sarah.elton@hyms.ac.uk

José Alexandre Felizola Diniz Filho Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiania, GO, Brasil, diniz@icb.ufg.br; jafdinizfilho@gmail.com

José Manuel Fuertes Departamento de Informática, Escuela Politécnica Superior Universidad de Jaén, Campus de las Lagunillas, 23071, Jaén, Spain, jmf@ujaen.es

Melanie Gibbs Ecology and Biogeography Unit, Biodiversity Research Centre, Catholic University of Louvain (UCL), 1348 Louvain-la-Neuve, Belgium, melanie_ri_gibbs@yahoo.co.uk

Øyvind Hammer Natural History Museum, University of Oslo, Oslo, Norway, oyvind.hammer@nhm.uio.no

Jonathan D. Krieger Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK j.krieger@nhm.ac.uk

Matthew D. Larvan Department of Earth Sciences, University of Bristol, Bristol, BS8 1RJ, UK, matt_larvan@hotmail.co.uk

Michael J. Lenardi Cultural Resource Survey Program, New York State Museum, Cultural Education Center, Albany, NY 12230, USA, mlenardi@mail.nysed.gov

Kay Lucek Department of Biology, Institute of Ecology & Evolution, University of Bern, CH3012, Bern, Switzerland, Kay.Lucek@eawag.ch

Manuel Jesús Lucena Departamento de Informática, Escuela Politécnica Superior Universidad de Jaén, Campus de las Lagunillas, 23071, Jaén, Spain, mlucena@ujaen.es

Ana L. Martínez-Carrillo Centro Andaluz de Arqueología Ibérica, Universidad de Jaén, España, Spain, anamartinezcarrillo@gmail.com

Jesús Marugán-Lobón Unidad de Paleontología, Dpto. Biología, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain jesus.marugan@uam.es

Thomas Merckx Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, Tubney, Abingdon, OX13 5QL, UK, thomas.merckx@zoo.ox.ac.uk

Daria E. Merwin Department of Anthropology, State University of New York at Stony Brook, Stony Brook, NY, USA, dmerwin@notes.cc.sunysb.edu

Richard A. Reyment Naturhistoriska Riksmuset, S10405 Stockholm, Sweden richard.reyment@nrm.se

Denis Roy EAWAG, Swiss Federal Institute for Aquatic Sciences and Technology, Center of Ecology, Evolution & Biogeochemistry, Kastanienbaum, Switzerland; Biology Department, Dalhousie University, Halifax, NS B3H 4J1, Canada, denisroy1@gmail.com

Arturo Ruiz Centro Andaluz de Arqueología Ibérica, Universidad de Jaén, España, Spain, arruiz@ujaen.es

Ole Seehausen EAWAG, Swiss Federal Institute for Aquatic Sciences and Technology, Center of Ecology, Evolution & Biogeochemistry, CH6047, Kastanienbaum, Switzerland; Department of Biology, Institute of Ecology & Evolution, University of Bern, CH3012, Bern, Switzerland, ole.seehausen@eawag.ch

Richard E. Strauss Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131, USA, Rich.Strauss@ttu.edu

Stefan Van Dongen Group of Evolutionary Biology, Department of Biology, University of Antwerp, B-2020 Antwerp, Belgium, stefan.vandongen@ua.ac.be

Hans Van Dyck Ecology and Biogeography Unit, Biodiversity Research Centre, Catholic University of Louvain (UCL), 1348 Louvain-la-Neuve, Belgium, hans.vandyck@uclouvain.be

Mark T. Young Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK; Department of Palaeontology, Natural History Museum, Cromwell Road, London, SW7 5BD, UK, mark.young@bristol.ac.uk

Part I
Introduction and Historical Review of the Subject (Chaps. 1 and 2)

Chapter 1
Why Morphometrics?

Ashraf M.T. Elewa

Five years since Springer published my first book on morphometrics (Elewa 2004), this short interval has shown rapid evolution in the subject.

At this point, I should note that the review made by Dr. Zelditsch (http://www.nhm.ac.uk/hosted_sites/pe/2005_2/books/morpho.htm) on the first book of 2004 is not convincing to me, despite it contains positive points, for the following reasons:

1. I wrote in the introduction of the book, as well as Prof. Rohlf did in the preface, that this book is principally oriented to systematists. Therefore, most applications were focused on systematics. Consequently, the stress on that the book did not involve some other branches is out of place.
2. One of the successful aims of the book was to show beginners how to use different techniques of morphometrics for solving different problems in systematics.
3. Another aim was to invite biologists and paleontologists to work together for solving their taxonomic problems in a compatible manner.
4. Proudly, the book was one of the earliest books to introduce examples of 3-D morphometrics (Chap. 7 of Kaandorp and Garcia Leiva; Chap. 17 of Harvati).
5. All over all, the book is considered by **NHBS** Environment Bookstore (http://www.nhbs.com/bioinformatics_cat_477-530-.html) as one of the best and most popular books related to bioinformatics on Earth. It is arranged at higher rank than many well known books on morphometrics. Also, it is considered by many famous bookstores as one of the bestselling books worldwide.

Anyway, morphometrics as a science attracted several students and researchers with broad interests due to the urgent need for quantification and visualization of shape to solve numerous problems related to wide ranges of scientific research.

———————————

A.M.T. Elewa (✉)
Geology Department, Faculty of Science, Minia University, Box 61519, Minia, Egypt
e-mail: aelewa@link.net; ashrafelewa@ymail.com

Earlier, in the seventieth of the last century (1971), when Blackith and Reyment published their book on multivariate morphometrics, this subject was new and applications were almost limited to solve biological problems. Thereafter, morphometrics became known for some other branches (e.g. geology and behavioral sciences).

In the new century, many other branches of science have been deeply involved (e.g. medicine, geomorphology, anthropology, art, and even resolving criminal mysteries through analyzing shape of footprint and shoes print).

Another progress of this subject is clearly related to the use of 3-D moprhometrics (see Figs. 1.1, 1.2, and 1.3 by Rob O'Neil). Although 3-D morphometrics is in its earlier stages, however I remember the words of Prof. Rohlf, in the preface of my first book, who said then that the next few years we are likely to see very exciting new developments that should add even more power to morphometric studies in systematics.

Prof. Norman MacLeod, who is one of the famous experts in the field, stated, in his 2005 review of the book titled "Geometric Morphometrics for Biologists – A Primer", that the authors set out to write such a book for the audience best positioned to appreciate a primer: those with no more than undergraduate mathematical training who want an emphasis on applications rather than theory. He added that there is certain need for such treatment. He also indicated that canonical modern morphometric texts are too technical and abstract to be fully understood by those not interested in making a commitment to the mathematics. Simultaneously, he said, collections of generalized applications articles are too eclectic and lack the unified focus necessary

Fig. 1.1 Data-projection by Rob O'Neil
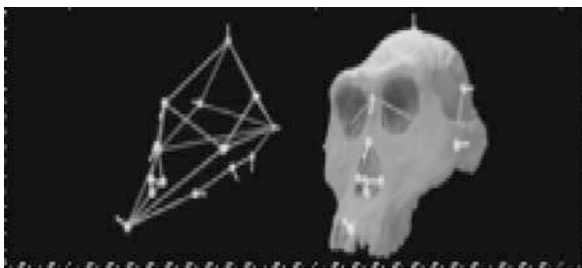
Fig. 1.2 Data-projection by Rob O'Neil

Fig. 1.3 Data-projection by
Rob O'Neil

original   generated
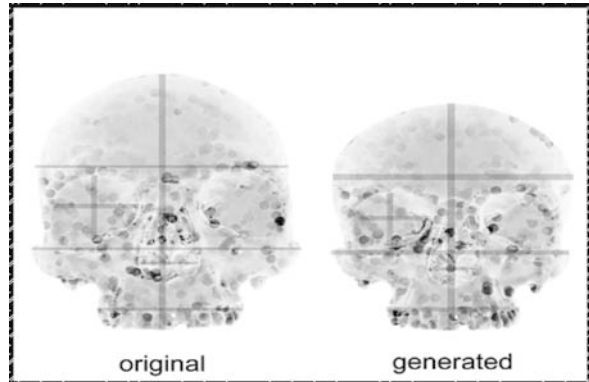
to be used as comprehensive introductions, while the special-topics collections are
too focused and contain too much nonmorphometric material. At the end of the
paragraph Prof. Macleod asked – Is this the primer we've all been waiting for?

Keeping this in mind, the aim of the present book is introduced in the fol-
lowing lines through explaining my philosophy of editing two books related to
morphometrics within short interval.

Essentially, morphometricians should have basic knowledge of mathematics and
statistics to understand how they can solve their problems with suitable and effec-
tive techniques. In fact, we all know that many colleagues are not specialized in
mathematics. Therefore, the goal of editing this book lies in two points:

1. To facilitate the understanding of morphometrics through introducing simple
   methodological approaches with examples. Actually, many books dealing with
   morphometrics appeared since the pioneer work of Prof. Bookstein (1989, 1991).
   These books (e.g. Marcus et al. 1993; Rohlf and Bookstein 1990; Marcus et al.
   1996; Small 1996; Dryden and Mardia 1998; Macleod and Forey 2002; Zelditsch
   et al. 2004; Julien 2008) are almost oriented to morphometricians with good
   knowledge of mathematics.
2. Confirming the effectiveness of morphometrics in resolving problems related to
   broad interests.

To come to the point, this book represents an up to date extension to my first
book on morphometrics that was published in the year 2004 with Springer-Verlag.
The previous book was successful to represent to readers several applications of
morphometrics in the fields of biology and paleontology. Nonetheless, we should
think of books that could attract more readers on the subject through publishing
books with simple scientific issues and modern techniques as well as up to date
information on the subject matter with examples. That is why I aimed to publish the
book in hand.

Since the seventieth of the last century (1971), when Blackith and Reyment introduced their book on multivariate morphometrics, there were no books focused on morphometrics for nonmorphometricians. This book is aimed to introduce the subject in its simplest form, keeping in mind that the readers have poor knowledge of mathematics and statistics. Yet, we cannot neglect the progress of this branch of science through introducing new ideas with modern techniques and applications.

To achieve the goal, I arranged the book in the form of textbook style with input of well known morphomometricians. So, we can have the two advantages assembled together; the style of textbook and the experience of several experts in the subject. It is not my way this time (it is different from what I made in my first book on morphometrics) to assemble papers representing applications of different groups of organisms. Therefore, the book is divided into four divisions:

1. Part one: Introduction and historical review of the subject.
2. Part two: Methodological approach (quantification of shape, and traditional and geometric morphometrics).
3. Part three: Applications.
4. Part four: Scope to the future of morphometrics.

Accordingly, the book includes, at first, four reviewing and methodological chapters with examples, adding to this introduction (Chap. 2, by Reyment, as a historical review of the subject; Chap. 3, by Krieger, on quantification of shape; Chap. 4, by Strauss, on discriminating groups of organisms; and Chap. 5, by Barcelo, on visualizing archaeological data). Then, there will be different chapters focusing on applications. Finally, an interesting chapter, written by Strauss, on the future of morphometrics is included, followed by a short note on morphometrics and cosmology by Elewa, and ended with a worthy of note chapter by Hammer on the advantage of integrating morphometrics with both general statistical and special ecological analysis modules in the same program, with special reference to the PAST software Package (Hammer et al. 2001).

Following my edited books with Springer, I invited an exceptional group of specialists to write on the subject. Therefore, I would like to thank them all, and particularly Prof. Rich Strauss (USA) for writing the foreword as well as two chapters for the book. The publishers of Springer-Verlag are sincerely acknowledged. I am also grateful to the staff members of Minia University of Egypt. Rob O'Neill, the acting director and research associate of the Digital Arts Research Laboratory, Pratt Institute, USA, is deeply acknowledged for providing me with some amazing pictures of his own work to use in this book.

Certainly, morphometrics became one of the most popular topics for students, at all levels, researchers, and professionals. Accordingly, this book represents advanced ideas and useful outline, especially for undergraduate and postgraduate students.

Now, if you are interested to know more about "why morphometrics?", then go through the chapters.

References

Blackith RE, Reyment RA (1971) Multivariate morphometrics. Academic Press, London, 412 pp

Bookstein FL (1989) Principal warps: thin-plate spline and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11: 567–585

Bookstein FL (1991) Morphometric tools for landmark data: geometry and biology. Cambridge University Press, New York, 435 pp

Dryden IL, Mardia KV (1998) Statistical shape analysis. John Wiley & Sons, New York, 347 pp

Elewa AMT (2004) Morphometrics – applications in biology and paleontology. Springer-Verlag, Heidelberg, 263 pp

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4 (1):9

Julien, C (2008) Morphometrics with R. Springer, New York, 316 pp

MacLeod N (2005) Review of the book titled "Geometric Morphometrics for Biologists – A Primer". The Palaeontological Association Newsletter 58:72–78

MacLeod N, Forey PL (2002) Morphology, shape and phylogeny. Taylor and Francis, London, 304 pp

Marcus L, Bello E, Garcia-Valdecasas A (1993) Contributions to morphometrics. Madrid, Monografias del Museo Nacional de Ciencias Naturale Madrid, 264 pp

Marcus L, Corti M, Loy A, Slice D (1996) Advances in morphometrics. Plenum Press, New York, 587 pp

Rohlf FJ, Bookstein FL (1990) Proceedings of the Michigan morphometrics workshop. Special Publication 2, The University of Michigan Museum of Zoology, 380 pp

Small CG (1996) The statistical theory of shape. Springer, New York, 227 pp

Zelditsch ML, Swiderski DL, Sheets HD, Fink WL (2004) Geometric morphometrics for biologists, a primer. Elsevier Academic Press, New York, 443 pp

Chapter 2
Morphometrics: An Historical Essay

Richard A. Reyment

Idea and Aims

The aim of this chapter is to inform morphometricians at large on the historical background of MORPHOMETRICS and to present the work and thoughts of the originators of the subject, with emphasis on the pioneering achievements of Professor Robert E. Blackith. Links between the latest developments of an aspect of morphometrics, in which geometric aspects are stressed, are outlined.

Introduction

The concept of morphometrics has a long history, notwithstanding that most of what is housed under its aegis is of quite recent origin. In this essay an outline of the two faces of morphometrics is attempted. The emphasis is on the historical development of the branch of knowledge over time, the orientation being one of respecting a state of intellectual achievement *à l'epoque* and without pointing the finger or scorn at past work considered proven by current thought as being misguided or inadequate. Reyment (2005) is a more detailed review of applied morphometrics in thought and praxis to which the reader may be referred for notes on palaeontological applications.

Morphometrics may be defined as a more or less interwoven set of largely statistical procedures for analysing variability in size and shape of organs and organisms. Some of the concepts have been generalized to encompass non-biological problems. Such areas are not taken up in the following. For a complete account of the subject of generalized shape as a mathematical concept, the most authoritative reference is Kendall et al. (1999). Attempts at expressing variability in size and shape in quantitative terms have a relatively long history in biology. The basic principle of what came to be known as principal component analysis was also presented

R.A. Reyment (✉)
Naturhistoriska Riksmuset, S10405, Stockholm, Sweden
e-mail: richard.reyment@nrm.se

in a shape-size context. However, its development and application were held back owing to computational difficulties. Some noteworthy biological studies did, however, emerge for material based on just a few variables, for example, the work of Pearce (1959, 1965) on the expression of shape in fruit-trees.

The term "Morphometrics" seems to have been coined by Robert E. Blackith (Professor of Zoology at University College, Dublin University) some 50 years ago (Blackith 1957), who entered into the subject through his engagement with the agricultural problem caused by swarming locusts. He applied multivariate statistical methods to the basic carapace morphology of grasshoppers and was able to follow the likelihood of the development of a swarming phase in a population by pinpointing morphological changes heralding a population explosion. This approach is clearly biological and Blackith's results represent one way of introducing a precise biological model into a statistical analysis of traditional stamp.

Essential to morphometric analysis is the requirement that variation in shape be represented in describable and repeatable terms. The roots of the geometric aspects of morphological graphology is usually conceded to lie with the work of the 15th century artist Albrecht Dürer, usually said to be German but who in actual fact was Hungarian. Dürer's father relocated to Nüremberg where he changed his name first to Thürer (from the Hungarian Ajtósi, doormaker), then to Dürer, more in keeping with the local dialectal pronunciation of Thürer. In geometrical terms, Dürer used the properties of affine transformations for distorting details (such as faces) by lateral or vertical elongation, thus mapping a square into a parallelogram and, or, by shearing a feature.

Blackith (1965) pointed out that the earliest recorded attempts to compare the shapes of animals was made by the school around Pythagoras as early as the 5th century BC. A rough drawing of a plant or animal that recorded the essence of the organism by noting the number of junctions between the lines of the sketch, but with no more than the minimum number of lines for representing the image of the organism provided the basis of the procedure.

Somewhat later it seems, the ancient Egyptians were concerned with embellishing burial monuments with figures and scenes, carved in limestone. Examination of photographs of some of these works of art, for example those figured in the treatise by Gay Robins (1994) discloses that some of them retain a discernible pattern of standardized squares, marked in red "chalk" which form a framework for making the carvings. It is apparent that the finished work of art was meant to be cleansed of the "props", but this did not always take place. How were the squares useful? She, and others, have noted that there was a reigning system of conventions, presumably determined by a collegium of experts, which defined the proportions of the human body to be used in adorning graves and memorials. The proportions of the limbs were standardized to a given number of squares, or a part of a square. It is only in the details of the head that a genuine likeness could be introduced. The square-standards were maintained for hundreds of years, then changed for a lengthy period, only to be brought back to the original conventions, probably by decree.

The skilfully stylized system of the ancient Egyptians for representing human proportions was to appear many hundreds of years later in the hands of Dürer and

da Vinci, and at a time when the system of rules of several thousand years earlier had neither been understood nor, even, suspected. Dürer's "menschliche Proportionen" were little more than the standardized Egyptian squares, though now expressed as lengths in terms of fractions of total height and, at times, affinely transformed. The well known figure of a man inscribed in a circle attributed to Leonardo da Vinci is a demonstration of the same idea, more succinctly assessed by Dürer (1512, 1513, 1520).

The origins of the quantitative analysis of measurements on organisms lie with the early biometricians, Karl Pearson (1857–1936), Francis Galton (1822–1911) and W. F. R. Weldon (1860–1906). A review of the pioneer studies of this triumvirate, and others), is given in (Reyment 1996). One of the methods, the correlation coefficient, invented by Galton and mathematically anchored by Pearson, is a standard tool of biometric analysis. It is worth considering why Pearson did not make more of his finding that correlations computed between proportions are not valid – he called them spurious correlations. The reason may well be, as has been surmised, that Pearson was not really interested in becoming entangled in biological conjecture, his main interests lying elsewhere. Nonetheless, two years before his death, Pearson and Morant (1934) published a detailed biometrical analysis of the "Wilkinson skull", proving beyond doubt that it was indeed the mummified skull of Oliver Cromwell, left to its fate on a pole after a gruesome post-mortem beheading.

Morphometrics as an Algebraic Exercise

As effective computational facilities became more generally available, so did interest in studying shape variation increase. A rather simple procedure rapidly gained ascendancy that relies on interpreting the latent roots and vectors of a matrix of covariances or correlations in terms size and shape components. The idea seems to have been formulated by the French marine biologist Teissier (1938) in his work on crabs (but only for the first principal component) and then improved and introduced into the anglophone literature by Jolicoeur and Mosimann (1960) and later provided with a more logical mathematical model by Hopkins (1966) using principal component factor analysis. Jolicoeur (1963) developed his original idea further by successfully relating the latent vector reification of principal component analysis to an allometric model. Although the general "algebraic solution" is usually attributed to Jolicoeur and Mosimann (1960), Quenouille (1952) was, it seems, probably the first worker in the field to give latent vectors a biological shape-oriented interpretation, closely followed by Pearce, who summarized his work on the growth of trees in a statistical textbook (Pearce, 1965). Sprent (1972) proposed a formalization of the prevailing concepts at the time for the analysis of size and shape.

The principal component method for describing size and shape variation relies on an intrinsic property of non-negative matrices known as the Perron-Frobenius theorem which states that the maximum root of the matrix is associated with a latent vector with positive components. The first latent vector of a covariance or correlation matrix of distance-measures observed on some organism is interpreted

as indicating variation in size. Subsequent latent vectors are said to be indicative of various factors of shape-variation. This interpretation is circumstantial in that it is in part based on an artefact. Nevertheless, the method is still in wide use and there is some evidence that supports the claims for a useful spectral decomposition of variability in size and shape. The same methodology is applied in Geology to the study of sedimentary data, species compositions in palaeoecology, and in the analysis of geochemical data. In these connexions there may be more convincing justification for the interpretation (reification) of the latent vectors. The results of multivariate analyses of distances observed on fossils are often presented as ordinations. That is, as bivariate scatter plots of the scores obtained from substituting the columns of the data-matrix into the latent vectors.

Growth and shape-change include a tensorial component in that they are at different rates in different directions at typical locations in a tissue. Huxley (1932) was well aware of this distinction. This directionality is not easy to test in a principal component decomposition, but easy of access if a geometrical orientation is adopted.

Mosimann's (1970) paper on allometry, and the "identification" of the size vector can be said to mark the starting point for a more geometrical approach to morphometrics and one which lies near to the heart of the new geometric morphometrics but without going all the way.

Blackith (1957) embarked upon an ambitious programme aimed at establishing a quantitative basis for studying growth and form in insects. The initial publication can be seen as an attempt at mapping out a program for detailed research with standard methods of multivariate analysis such as were available at the time, in a biological vista and with emphasis on the results being achieved by the Indian school (Calcutta) in the hands of P. C. Mahalanobis and C. R. Rao (1952) with reference to the anthropometrical survey of the cast- system in India (United Provinces). The procedures reviewed in that work were the linear discriminant function, the method of multiple discrimination (canonical variate analysis) and principal components. A feature of this short article is the use made of a chart of generalized distances along two axes, one for phase and one for size. Within the field of experimental biology, Blackith (1957, 1960), in a suite of papers devoted to polymorphism in insects, in which Mahanolobis' distances and canonical variates were handled with great insight, established a foundation for the quantitative study of variation in size and shape. A major problem to be solved in using "distance measures" in the study of variation lies with the fact that size and shape are confounded. This means that a suite of measurements on a locust carapace, for instance, represent not only a component expressing size-differences but also differentiation between variables that arises from differences in shape of the specimens of a sample. Blackith could reduce the negative aspect of this by a clever graphical method of linking his distances in relation to known polymorphisms. This could not however hope to provide a general, unequivocal solution to the problem of the analysis of shape.

Blackith et al. (1963, p. 317) observed that which has also been done by others that Thompson's (1917) application of coordinate transformations was done in a semi-quantitative manner and never expanded and could not take account of (nor

indeed was envisaged as doing) such a fundamental attribute as that of differential growth and the change or relative proportions with absolute size. A solution to that problem came first with Huxley (1932) in his celebrated volume on relative growth (N.B. Huxley did not use the term "allometry" in his treatise of 1932 for relative growth – his term was heterogeny).

The Q-mode method of principal coordinate analysis has shown itself to be useful for presenting graphical analyses of relationships such as specific differences and intraspecific differences such as polymorphism in characters, not only morphological features but also features that are dichotomous and qualitative. Gower (1966) developed this technique as an answer to the inappropriate analysis known as Q-mode factor analysis, a procedure which, however, is neither an authentic factor model nor constructed for preserving distances between specimens (Reyment and Jöreskog 1999). Principal coordinate analysis is, moreover, not inverted principal components. In order for the technique to function in a correct manner, Gower's (1971) similarity matrix is a necessary prerequisite. This use of this technique permits the user to combine quantitative, binomial and qualitative attributes in the same distance-preserving representation. In the special case of a matrix of correlations of correlation coefficients, principal components is the R-mode dual of Q-mode coordinates, and the necessary manipulations can be simply performed by Sylvester's extraction of latent roots and vectors – now generally known as the singular value derived by decomposition of a square symmetric matrix in the more general Eckart-Young solution. Be it noted however that this is a restrictive procedure in that it cannot encompass dichotomous and qualitative variables and hence is of limited value in very many studies, and particularly in palaeontology where the number of characters may be diverse as well as numerous.

How biologically sound are distance-based characters in the hands of mathematically oriented biologists? This is a question raised occasionally by the more geometrically schooled practitioners. My experience is that the selection of "taxonomically relevant distances", which are, in effect, distances between "landmarks", is usually done in a very conscientious manner and well anchored in a detailed knowledge of the biology of the species under study. The proof of the pudding is in the eating, as it were, and in the case of entomological work on harmful insects, it is been found by experience that distance measures, judiciously selected, yield valid, intelligible results. The term "landmark" is not a great terminological borrowing. Its invasion of the biometrical literature comes via its applications in craniology, the study of variation in the dimensions of skulls of primates, and basically a subject not of prime interest to the mathematician. The real meaning of the word is, as any anglophone knows, (1) object marking boundary of country, estate, etc., (2) conspicuous object in a district, (3), object or event or change marking stage or process or turning-point in history.

Before leaving the insects, it is interesting to note that Blackith and Kevan (1967, p. 81) took up the geometrical problem of the space in which canonical variates are located. Campbell (1979, Canonical variate analysis: some practical aspects. PhD Thesis, unpublished) has made an extensive study of the structure of the canonical variate model. He showed that it is not to be expected that the vectors of canonical

variate analysis are directly equatable with the latent roots of principal component analysis. This is often a matter of confusion to the mathematically untrained users of computer programs.

Morphometrics as Interpreted by Statistical Mathematicians

The English mathematician Maurice S. Bartlett, who was intimately connected with the development of several major areas of mathematical statistics, also made contributions to the application of multivariate statistics to quantitative biology (Bartlett, 1965). Admirable though as this is as to the clear presentation of multivariate methodology, Bartlett's insight into biological processes leaves much to be desired.

T. P. Burnaby began his professional life as a palaeontologist at the University of Keele, U. K. His studies were from the outset centred around quantitative aspects of various invertebrates, mainly foraminifers and bivalves. Burnaby died prematurely (1924–1968) and his widow deposited his research notes with me at the University of Uppsala in 1970. These notes show Burnaby to have been gifted with mathematical acuity and it is clear that at the time of his passing he was concerned with several biological problems which, if they had been pursued to fruition, could well have led to major contributions in the analysis of growth and form. Burnaby was much concerned with arriving at a reliable procedure for studying size and shape in bivalves. Bivalves do not have a terminal growth-size and grow additively, which implies that just keep getting larger and larger right unto the end. Burnaby observed that the growth pattern was not a mere regular amplification of the shell but that there were steps that were lacking in perceptible regularity. He made a similar observation for the growth of coiled ammonite shells (genus *Cadoceras)* which on the basis of the Reverend Mosely's identification of the logarithmic growth spiral for ammonites, were considered to develop in a regular manner, but which did have a terminal growth-size. Burnaby's notes, and Burnaby (1966a), show that he had found that the angle of growth (the spiral angle) of presumably regularly coiled ammonite shells is not constant but can, and does, fluctuate during growth.

Thus, the real nature of the problem of confounding of size and shape during growth was first realized clearly by Burnaby (1966b) who devised a mathematical procedure for a transformation that placed "size" in one subspace and shape into another. He named his method "growth-free discrimination". This solution is not often used. It is nonetheless a remarkable and foresighted achievement that doubtlessly influenced directly or indirectly the current phase of development of the subject. Burnaby's (1966b) quest for a discriminant function that would override the effects of confounding due to size differences, polymorphism, ecologically stimulated effects (ecophenotypy), etc., representable as gradients, had a definite circumscribed objective. He was definitely not concerned with describing shape variability itself, as is sometimes assumed, but solely with addressing a biostratigraphical problem occurring with organisms that do not have a terminal growth size. Burnaby was fortunate in having C. Radakrishna Rao as one of the referees for his

submission to *Biometrics* and for Rao being willing to help provide the text with a rigid mathematical framework (Burnaby's correspondence). A major problem at the time was that Burnaby was unable to find real data for exemplifying his procedure and was obliged to use an artificial example which almost led to the rejection of the submission (correspondence in Burnaby's manuscripts). Rao (1966) moved quickly to produce a theoretical paper based on his knowledge of Burnaby's submission and in part inspired by it. However, Rao concentrated on discrimination aspects and hence on two populations, but did not discuss problems of estimation.

Gower (1976) took up the algebra of Burnaby's growth invariant discriminant functions in a more detailed manner. Gower noted that the conceptual basis of the original work of Burnaby was not unchallengeable, not least because of the problem associated with the estimation of linear growth effects and a difficulty which Burnaby had not been able to surmount. Reyment and Banfield (1976) applied Gower's panoply of methods for growth-adjusted canonical variates to species of Paleocene foraminifers. Among the methods proposed by Gower for estimating growth etc. vectors, we used two of them, to wit, the principal component solution and approximation by maximum likelihood factor analysis (i.e. "True factor analysis" (Reyment and Jöreskog 1999). The principal aim of our analyses was to use the canonical variate means for comparisons between species observed at different time-intervals and not to attempt an analysis of shape variability. This is an important limitation on the scope of the study and one worth keeping in mind. The results obtained for the methods proposed by Gower (1976) showed that the principal components of the pooled within-samples covariance matrix were successful in removing the major source of conflicting variation. This variation was found to derive from individuals of the foraminiferal species having been at different stages of growth when fossilized. We also found that the patterns produced when the growth-free canonical variate sample means were plotted against chronological order led to useful evolutionary comparisons between species.

Supplemental Reading

The book by the plant-ecologist P. Greig-Smith (1957) is essential reading for anybody seriously interested in applying quantitative methods to the study of plants not least because of the detailed practical treatment of principal components and principal component factor analysis.

A becharming advanced introduction to biometric analysis is the book by C. R. Rao (1952), notwithstanding its almost 60 years since the first printing. Questions being asked by tyros today with respect to aspects of canonical variate analysis (Rao does not use this term in his book, preferring the designation multiple discrimination) are taken up in, for example, the case of Miss M. M. Barnard's (1935) Egyptian skulls and Fishers statistical treatment of her problem. And, as an intellectual exercise, the astute student is invited to find out what is wrong with Barnard's model from the biological aspect.

A mathematically rigorous treatment of what is known as "factor analysis" in many connexions is the text by Reyment and Jöreskog 1999. The distinction brought to the fore here is that the true factor model is seldom, if ever, encountered in biology. A surrogate of somewhat rickety validity is employed instead, the appropriate term for which is "principal component factor analysis". The confusion seems to have arisen with Teissier's (1938) expression "analyse factorielle" for a principal component analysis of crabs and a too literal translation, of the French text.

Reyment (1996) gave an historical review of the early background of morphometrics in which the evolution in thought on that subject is outlined.

The Coefficient of Racial Likeness was introduced into anthropometry by Karl Pearson. It was, however, nothing more than the generalized statistical distance of P. C. Mahanalobis (shown to Pearson in 1927 by Mahanalobis), with all off diagonal elements zero and variances along the diagonal. Pearson did not want to admit the superiority and relevance of the generalized distance, thereby embarking upon a fruitless controversy (cf. Mahanolobis, 1936).

Morphometrics in a Geometric Setting

The problem of how to construct a practical and mathematically justifiable solution to the coordinate-based concept of Thompson (1917, 1942) for geometrically illustrating shape relationships in organisms with the emphasis placed on phylogenetic reconstructions dogged biometricians for many decades. Thompson thought it should be possible to represent relationships between related organisms by deformational grids. By means of an affine transformation, one genus of fishes, for example, could be transformed graphically to another genus, hopefully phylogenetically connected. For years workers puzzled over how this was actually done – what did the mathematics look like? Speculation went on for years and years without much light being cast on the subject. Bookstein (1991) and Dryden and Mardia (1998, p. 200) noted the subjectivity attaching to Thompson's freehand transformations. Huxley (1932, pp. 104–110) discussed weaknesses he considered to be inherent in the method of Cartesian coordinates from the aspect of the constancy of growthgradients. Thompson's book was reprinted many times after 1917 but he did not in any of the editions refer to how he had made his figures. In any event, Thompson's insight was a major break-through in thought about biological variability but it was not until 1978 that Bookstein provided a solution for the affine case and then for the non-affine case (Bookstein 1986) in many later publications, summarized in Bookstein (1989, 1991). These works lie centrally located at the origin of the modern development of geometric morphometrics. It is no exaggeration to claim that without Bookstein's exceptional insight, and didactic skills, the geometric analysis of shape would not have developed into a biologically relevant discipline, capable of being expanded in many directions.

Measuring Outline Shape

Distance measures can be arbitrarily constructed around an outline by a simple geometrical method by marking off equispaced points. A well known means of describing a curve, such as an outline, is by decomposing the line into a Fourier Series which can be made to approximate the contour of the object by passing through a series of progressively more complex trigonometric functions for the digitized points. Although very good approximations of the shape of an object can be made, the results cannot be linked to homologous relationships between objects. The use of Fourier Series is well known to geologists from the sphere of analytical sedimentology.

Lohmann (1983) applied a result of Zahn and Rosskies (1972) to the study of shape in planktic foraminifers. This result shows that a complex curve can be expressed as a series of steps around a circle. The underlying mathematical concept of Lohmann's "eigenshapes" is conceptually similar to that of Burnaby's decomposition into size and shape spaces. In eigenshapes, size is represented by the length of the steps around the perimeter of the object and shape by the set of angles estimating the deviation of each step from the expected direction. Lohmann's work, considered in the context of time and knowledge, is an outstanding achievement. Swiderski et al. (2002) have given a thoughtful appraisal of eigenshape analysis. Macleod's (2002) extended eigenshape analysis is a serious attempt at coming to grips with the problem of comparability between objects. Bookstein (1997) proposed powerful coordinate-based methods for studying forms without landmarks, that is, for accessing the information in curving outlines.

A reasonable case for outline methods can be made for planispirally coiled shells such as those of ammonites. Natural, circumferentially located reference points are not available on such shells, even where ornamental features occur, owing to the inherent instability in such properties with respect to the number per whorl and the degree of development. Ammonites are often richly ornamented with tubercles and ribs. These features are, however, seldom stable enough to permit using them as a base for homologous landmark points. Greater availability is offered by the apertural aspect in which points of intersection of features occur. Here again the canon of geometric morphometrics is not always useful and for many studies, the analyst must perforce fall back on the multivariate analysis of distance measures (Dryden and Mardia, 1998).

Reference Points (Landmarks)

Geometrically based morphometrics is in its current form dependent on the selection of reference points, designated by X-, Y-coordinates and conveniently referred to as "landmarks". MacLeod (2002) has drawn attention to this and, for example, pointed out that taxonomic distances in current use are no more than measures between reference points on an object which in turn are just "landmarks". The same distinction

has been made by Dryden and Mardia (1998) and Bookstein (1991). Kendall et al. (1999, p. 1) espouse a rather rigid interpretation of labelled points (they adroitly bypass the use of the word "landmark") in their mathematical concept of markers in that they underline that for them, labelled points are basic and determine the objects studied. According to the biologically oriented concept of the "geometric morphometricians" the 'marker points' are selected from a usually two-dimensional or three-dimensional continuum. The biological interest does not encompass cases where markers all lie in lower-dimensional subspace or two or more of them coincide, which contrasts with Kendall's spaces, which contain the shapes of all possible configurations except those for which all the points coincide (Kendall et al. 1999, p. 2).

Landmarks are specified by pairs of X-, Y-coordinates (the usage is originally a borrowing by osteologists from topographical surveying, where fixes are located with respect to coordinate pairs). Using simple geometry, distances can be constructed from taxonomic reference points: the reverse procedure is, however, not possible. The arbitrary points on a circumference in eigenshapes are likewise landmarks, denoted by X-, Y-coordinates, but they lack the property of homology or point-to-point correspondence from specimen to specimen.

Thin-Plate Spline-Based Morphometrics

The concept of thin-plate splines for expressing, graphically, deformations is not based on a biological concept, in common with the algebraic solution. It seems to have been deveoped in connexion with French engineering work, possibly connected with stability in fuselages of ultrasonic aircraft . One must be sure of what the spline-based morphometrics can do. Evangelical claims for biological relevance made especially in the earlier stages of applications have not always been useful and, as just stated, not true. The selection of landmarks determines the outcome of the shape-analytical conclusions owing to the fact that the thin-plate spline decomposition is not rooted in covariation in shape-changes in the input-data. Hence, each configuration of landmarks is a unique representation of just the set of points selected on the object (Rohlf, 2002, p. 179). Deletion or addition of points is often found to change the visualization diagram and hence the interpretation. In consequence hereof, it is desirable to make clear that the results of a particular analysis pertain specifically to a particular configuration of reference points on an organism. Biological interpretation must inevitably be made with expert insight (Rohlf, 1993).

Palaeontological material consists almost exclusively of fossilized hard parts, the shells encasing the tissues and organs of the animal. Material obtained from living organisms, for example, brains, can be related to more than raw morphology which implies that an analysis of an organ can be given a more biological interpretation than can be hoped for by a palaeontologist. For example Bookstein's (1991) work on schizophrenia. The study of shape-variability in fossilized hard parts must perforce be in terms of "deformations"; this limitation implies that exceptional care must be expended on choosing reference points that really mean something in a

palaeobiological context and which can be extrapolated from case to case. In many cases, apart from fossil mammals, it is seldom possible to relate shape-variation in shell details to the anatomy of the soft parts of an animal. It is tempting to view a sequence of spline relationships resulting from the latent-root decomposition as being a kind of microscope-ratchet which successively yields a scale of magnifications of the surface of a fossil as a function of the latent-roots. For different configurations, different impressions of the topology of the surface will be communicated. Hence, there is considerable obligation placed on the investigator to select reference points that are of scientific significance. Large latent roots correspond to latent vectors that describe small-scale features - the deformation of landmarks that are close together. Small latent roots correspond to latent vectors that describe large-scale deformational features.

There is an important field in invertebrate palaeontology, however, that can supply much useful information from morphometric appraisal of shell properties and shape. This is the subject of polymorphism and polyphenism. Interesting results have been obtained for ostracods and cephalopods, for example, Reyment and Kennedy (1990) for ammonites and Reyment and McKenzie (1993) for Ostracoda from southern Australia.

Remarks on Spline-Based Methods

There are two main openings available for charting differences in form by means of coordinates. One of these takes each form, superimposes it in relation to others, and then computes differences in terms of reference-point displacements relative to this registration. The second tack is concerned with describing differences in point configurations as deformations of a grid produced by mapping one form into the other and visualizing the shrinkings and stretchings that are generated by the procedure. The analysis of the registrations of the coordinates may be done in several ways. One may register to a common baseline by translating, rotating and scaling so that most points fit well, or register by minimizing the sum of squared differences between the equivalent landmarks of forms. This is usually referred to as generalized Procrustean fitting. The result is scaled by division by the centroid size.

Affine and Non-Affine Transformations

The concept of affine transformations seems to be due to the celebrated mathematicians Möbius and Euler. In Physics, the idea of affine transformation is known as an homogeneous deformation (Klein, 1925, p. 75). Decomposition of coordinate-based data by the thin-plate spline technique exhibits a close analogy with the well known decomposition by means of Fourier trigonometric series. The constant term in such a series is a global parameter and the trigonometric coefficients are local parameters at successively smaller scales (Dryden and Mardia, 1998, p. 199). Dryden and Mardia (1998, p. 286) also point out that the results yielded by distances and coordinates are often similar and the former cannot be dismissed out of hand. For small

variations, registration methods and distance methods lead to identical conclusions about shape. Unfortunately, there is no general rule available to support this belief.

Possible Sources of Error

Errors can and do arise in the multivariate analytical processing of data generated by geometric morphometric procedures. The intuitively attractive method of triangulating data, Bookstein Coordinates, whereby coordinates are registered on a common edge induces spurious correlations and consequently invalid covariance matrices (Dryden and Mardia, (1998, p. 173). This notwithstanding, the basic theorem that everything that can be achieved using shape-coordinates can also be realized by means of ratios of distances, as demonstrated in the original paper on shape coordinates (Bookstein, 1986). The possibility of applying the theory of compositional data analysis (Aitchison, 1986) to the problem has not yet been more than briefly considered. Another source of error lies with imperfections in the data. This is more likely to be a problem in palaeontological work where deviations from multivariate normality are a cause of misleading results. Campbell (1979, Canonical variate analysis: some practical aspects. PhD Thesis, unpublished) seems to be the first worker to have given serious attention to practical aspects of consistency and repeatability in multivariate analysis, in particular canonical variates and discriminant functions, but also principal components. Some of Campbells's results are summarized in Reyment et al. (1984) and Reyment (1991).

Supplementary Reading

R. H. Benson (1972): Benson, a specialist on fossil and extant ostracods, was concerned with features of shell morphology, explicitly the reticulate pattern which he believed was under the control of evolutionary and environmental factors. This theory is not unchallengeable but it does contain an element of bio-geometric interest. He devised a method of graphical pattern analysis to be used for defining elements of the reticulum in a lavishly illustrated monograph. Benson's seldom cited monograph features series of reticular silhouettes with homologous regions identified with exemplary insight. The data were not analysed statistically, but remained at the visual level. Arising from Benson's graphical analysis, Siegel and Benson (1982) proposed a method for comparing shapes.

N. A. Campbell's (1982) publications listed in the references. These publications treat the problem of achieving stable results in methods of multivariate analysis, primarily canonical variate analysis but also principal components, with emphasis on the stability of the elements of the vector components. This is a question of more than trivial significance when the results obtained for, say, relative warps are to be expressed graphically or to be inserted into a standard extension of multivariate

techniques. As far as I have been able to judge, this awareness has yet to make itself apparent in geometric morphometrics.

Bookstein's (1978) publication on interpreting affine transformations has made a lasting impression on me and one that impresses by the depth of understanding it contains. Every student of morphometrics should consult that thesis now and then. The generalization of the field to encompass affine and non-affine changes in shape by Bookstein (1986) is likewise a classic. Both papers mark the dawning of the new era for the study of shape and growth.

A very important advance in the analysis of shape-data is the paper by Mardia and Dryden (1989). This work may be seen as marking the beginning of the use of statistical models for shape studies. The term "Mardia-Dryden" is due to Kendall (1991).

Concluding Remarks

The brief account of morphometrics presented in the foregoing pages can do no more than highlight a few of the most important features and events of what is a complicated subject caught up in a phase of expansive development. The introduction of advanced geometrical thinking into statistics is not widespread and very few statistical textbooks take it into consideration. The level of mathematical theory involved in shape-theory is certainly beyond the reach and learning of many statisticians and certainly transcends the ability of geologists and biologists. The truth of this statement can be assessed by consulting the volume by Kendall et al. (1999). An interesting and promising development in biological shape analysis has been proposed by Bookstein (2000) by means of a method called "creases", being an allusion to the pinched features characterizing the associated spline diagrams. The method of crease analysis examines the effect of expansions forwards and backwards in time in phylogenetic reconstructions (Bookstein, 2000). As far as I am aware, the potential of the 'crease' has yet to be exploited, which may reflect a dearth of phylogenetic analyses at the present time.

The crease method also holds promise of being a means of developing the automated description of diagnostic contrasts between reference specimens of species, that is, image-based taxonomy.

References

Aitchison J (1986) The statistical analysis of compositional data, Chapman and Hall, London, xv, 416 pp
Barnard MM (1935) The secular variation of skull characters in four series of Egyptian skulls. Annals of Eugenics (London) 7: 89 pp

Bartlett MS (1965) Multivariate statistics. In Waterman TH, Horowitz HJ (Eds.) Theoretical and mathematical biology, Chapter 8, Blaisdell Publishing Company, Newyork, 201–223 pp

Benson, RH (1972) The *Bradleya* problem, with descriptions of two new psychospheric ostracode genera, *Agrenocythere* and *Poseidonamicus* (Ostracoda, Crustacea). Smithsonian Contributions to Paleobiology, 12, 138 pp

Blackith RE (1957) Polymorphism in some Australian locusts and grasshoppers. Biometrics 13: 183–196

Blackith RE (1960) A synthesis of multivariate techniques to distinguish patterns of growth in grasshoppers. Biometrics 16: 28–40

Blackith RE (1965) Morphometrics. In Waterman TH, Horowitz HJ (Eds.) Theoretical and Mathematical Biology, Chapter 9, Blaisdell Publishing Company, Newyork, 225–249 pp

Blackith RE and Kevan DK (1967) A study of the genus *Chrotogronus* (Orthoptera). Evolution 21, 76–84

Blackith RE, Davies RG, Moy EA (1963) A biometric analysis of development in *Dysdercus fasciatus* (Hemiptera: Pyrrhocaridae). Growth 27: 317–334

Bookstein FL (1978) The measurement of biological shape and shape change. Lecture Notes in Biomathematics.– No. 24, Springer Verlag, New York, 191 pp

Bookstein FL (1986) Size and shape spaces for landmark data in two dimensions. Statistical Science 1: 181–242

Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions Pattern Analysis and Machine Intelligence 11: 567–585

Bookstein FL (1991) Morphometric tools for landmark data. Cambridge University Press, Cambridge, xvii, 435 pp

Bookstein FL (1997) Landmark methods for forms without landmarks: localizing group differences in outline shape. Medical Image Analysis 1: 225–243

Bookstein FL (2000) Creases as local features of deformation grids. Medical Image Analysis 4: 93–110

Burnaby TP (1966a) Allometric growth of ammonite shells: a generalization of the logarithmic spiral. Nature 209: 904–906

Burnaby TP (1966b) Growth-invariant discriminant functions and generalized distances. Biometrics 22: 96–110

Campbell NA (1982). Robust procedures in multivariate analysis. II Robust canonical variate analysis. Applied Statistics 31: 1–8

Dürer, A (1512, 1513, 1525) Vier Bücher von menschlicher Proportion. Nuremberg, Hieronymus Formschnyder, 264 pp

Dryden IL, Mardia KV (1998) Statistical shape analysis. Wiley & Sons, New York, xviii + 347 pp

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrics 53: 12–38

Gower JC (1971) A general coefficient of similarity and some of its properties. Biometrics 27: 857–874

Gower JC (1976) Growth-free canonical variates and generalized distances. Bulletin of the Geological Institutions of the University of Uppsala. N.S. 7: 1–10

Greig-Smith P (1957) Quantitative plant ecology. Butterworth, London, 256 pp

Hopkins JW (1966) Some considerations in multivariate allometry. Biometrics 22:747–760

Huxley JS (1932) Problems of relative growth. Methuen, London, xix, 276 pp

Jolicoeur P (1963) The degree of generality of robustness in *Martes americana*. Growth 27: 1–27

Jolicoeur P, Mosimann JE (1960) Size and shape variation in the Painted Turtle, a principal component analysis. Growth 24: 339–354

Kendall DG (1991) The Mardia-Dryden distribution for triangles – a stochastic calculus approach. Journal of Applied Probability 28: 225–230

Kendall DG, Barden D, Carne TK, Le H (1999) Shape and shape theory. Wiley and Sons, Chichester, x, 306 pp

Klein F (1925) Elementarmathematik vom höheren Standpunkte aus. Band 2: Geometrie. J. Springer Verlag, Berlin, 302 pp

Lohmann GP (1983) Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. Mathematical Geology 15: 659–672

Macleod N (2002) Phylogenetic signals in morphometric data. In MacLeod N, Forey PL (Eds.) Morphology, shape and phylogeny; Systematics association special volume 64, Taylor and Francis, London, 100–138 pp

Mahanolobis PC (1936) On the generalized distance in statistics. Proceedings of National Institute Science, India 2: 49–55

Mosimann JE (1970) Size allometry: size and shape variables with characterizations of the lognormal and generalized gamma distributions. Journal of American. Statistics Association 65: 930–945

Mardia KV, Dryden IL (1989) The statistical analysis of shape data. Biometrika 76: 272–282

Pearce SC (1959) Some recent applications of multivariate analysis to data from fruit trees. Annual Reports East Malling Research Station 1958: 73–76

Pearce SC (1965) Biological statistics. McGraw-Hill, New York, xiii + 212 pp

Pearson K, Morant, GM (1934) The Wilkinson head of Oliver Cromwell and its relation to busts, masks and painted portraits. Biometrika 26: 1–116

Quenouille MH (1952) Associated measurements. Butterworths, London, x + 242 pp

Rao CR (1952) Advanced statistical methods in biometric research. Wiley, New York, 292 pp

Rao CR (1966) Discriminant function between composite hypotheses and related problems. Biometrika 53: 339–345

Reyment RA (1991) Multidimensional palaeobiology. (+ appendix by LF Marcus) Pergamon Press, Oxford, ix, 377 pp

Reyment RA (1996) An idiosyncratic history of early morphometrics. In Marcus LF et al. (Eds.) Advances in morphometrics, NATO ASI series: Series A, life sciences, volume 284, 15–22 pp

Reyment RA (2005) Aspects of applied morphometrics. Zeitschrift der deutschen geologischen Gesellschaft. 154: 263–274

Reyment RA, Banfield C (1976) Growth-free canonical variates applied to fossil foraminifers. Bulletin of the Geological Institutions of the University of Uppsala. N.S. 7: 11–21

Reyment RA, Jöreskog KG (1999) Applied factor analysis in the natural sciences. Cambridge University Press, Cambridge, 371 pp

Reyment, RA, Kennedy WJ (1990) Phenotypic plasticity in a Cretaceous ammonite analyzed by multivariate statistical methods. Evolutionary Biology 25: 411–426

Reyment RA, McKenzie KG (1993) Quantitative genetics in Paleontology: evolution in Tertiary Ostracoda. International Association for Mathematical Geology: studies in Mathematical Geology, Computers in Geology – 25 years of progress. In Davis JC, Herzfeld U (Eds.) University Press, Oxford, No. 13, 155–168 pp

Reyment RA, Blackith RE, Campbell NA (1984) Multivariate morphometrics, Second Edition, Academic Press, London, vii, 233 pp

Robins G (1994) Proportion and style in ancient Egyptian art. University of Texas Press, Austin, 283 pp

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. In Marcus LF et al. (Eds.) Contributions to morphometrics, Museo nacional de ciencias naturales, Madrid, 131–159 pp

Rohlf FJ (2002) Geometric morphometrics and phylogeny. In MacLeod N, Forey PL (Eds.) Morphology, shape and phylogeny, Systematics association special volume series 64, Taylor and Francis, London, 175–193 pp

Siegel AF, Benson RH (1982) A robust comparison of biological shapes. Biometrics, 38: 341–350

Sprent P (1972) The mathematics of size and shape. Biometrics, 28: 23–28

Swiderski DL, Zelditch ML, Fink WL (2002) Comparability, morphometrics and phylogenetic systematics. In MacLeod N, Forey PL (Eds.) Morphology, shape and phylogeny, Systematics special volume series 64, Taylor and Francis, London, 67–99 pp

Teissier G (1938) Un essai d'analyse factorielle. Les variants sexuels de *Maia squinada.* Biotypologie, 7: 73–96

Thompson DW (1942 [1917]) On growth and form. University Press, Cambridge, second edition, 1116 pp

Zahn CT, Roskies RZ (1972) Fourier descriptors for plane closed curves. IEEE Transactions on Computers, C-21: 269–281

Part II
Methodological Approach (Quantification and Visualization of Shape, and Traditional and Geometric Morphometrics) (Chaps. 3 to 6)

Chapter 3
Controlling for Curvature in the Quantification of Leaf Form

Jonathan D. Krieger

Idea and Aims

Curvature and asymmetry are ever present in biological datasets as sources of shape variation, potentially confounded with other measurements of shape. This chapter examines potential methods for minimizing the impact of curvature on leaf shape measurement. Specifically: can curvature be ignored, should a curvature reference be included in analyses, or is mathematical straightening required? Methods used to measure leaf shape are reviewed.

Introduction

Botanical morphological diversity starts with leaf shape. Leaves are present throughout the land plants, found in vast quantities in the fossil record, and easy to preserve relatively unchanged as pressed herbarium specimens. For at least three hundred years, botanists have found applications for qualitative descriptions of tip shape, base shape, overall shape, and innumerable other characters. The number of terms developed to qualitatively describe leaf morphology is astounding, and they vary depending on the group of plants being studied (ferns, conifers, flowering plants). For the small set of fern species used in this study, there were six terms for tip shape, seven for base shape, and eleven for overall shape (Fig. 3.1).

Despite their long history, qualitative descriptions have shortcomings that limit their usefulness. They rely on categories, for example, categorizing a leaf tip as "round" or "straight". However, leaves are notoriously difficult to consistently score for traits that may vary continuously (Wilf 1997; Wiemann et al. 1998). To describe the range of variation in leaf shape within a species, terms are combined into descriptions like "narrow lanceolate to oblanceolate, rarely elliptic." This sort of

J.D. Krieger (✉)
Department of Palaeontology, The Natural History Museum, Cromwell Road,
London, SW7 5BD, UK
e-mail: j.krieger@nhm.ac.uk

overall shape
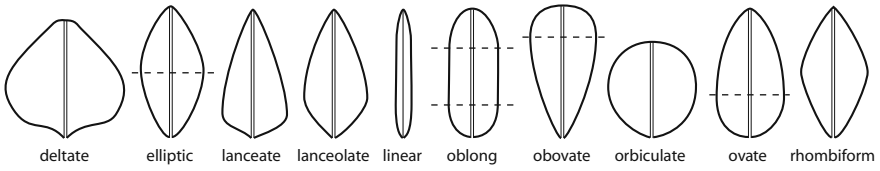
| deltate | elliptic | lanceate | lanceolate | linear | oblong | obovate | orbiculate | ovate | rhombiform |

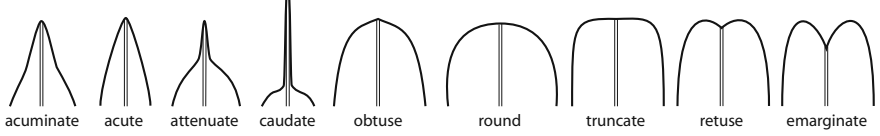|	
deltate:	broadly triangular, acute apex
elliptic:	oval, with the widest portion near the middle
lanceate:	long-tapering apex, nearly truncate base
lanceolate:	long-tapering apex, short-tapering base
linear:	long, narrow, of nearly uniform width
oblanceolate:	gradually tapering base, nearly truncate apex

oblong:	longer than wide, sides parallel, ends abrupt
obovate:	narrow base, round to obtuse apex
orbiculate:	round
ovate:	broadest toward base
rhombiform:	parallelogram-shaped, often longer than wide

tip shape

| acuminate | acute | attenuate | caudate | obtuse | round | truncate | retuse | emarginate |

acuminate:	less than 30°, the sides somewhat concave
acute:	30–90°, the sides straight or slightly convex
attenuate:	less than 15°, narrowly tapering
caudate:	bearing an elongate, tail-like apex
obtuse:	90° or more, sides straight to slightly convex

round:	circular rather than angular at the apex
truncate:	leaf terminates abruptly, as if cut
retuse:	midvein length is 95–99% total length
emarginate:	midvein length is 75–95% total length

base shape

| acuminate | acute | attenuate | cuneate | decurrent | excavate | truncate |

acuminate:	less than 30°, the sides somewhat concave
acute:	30–90°, the sides straight or slightly convex
attenuate:	less than 15°, narrowly tapering
cuneate:	wedge-shaped, 30–45°, the sides straight
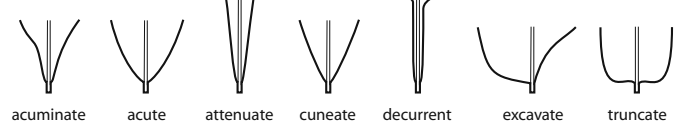
decurrent:	extending basiscopically along an axis
excavate:	having the basiscopic base cut away
truncate:	appearing as if cut off perpendicular to the axis

Fig. 3.1 Fern leaf shape terminology. Terms apply to the lamina of simple leaves and to the blade for dissected leaves. All terms shown have been applied to the species included in this study. Definitions modified and figures redrawn from Lellinger (2002)

amalgam of terminology can be used to describe subtle differences in shape, but the volume of available terms makes the descriptions hard to reproduce between investigators; it is seldom clear where the boundaries among categories are drawn. And, illustrating their lack of concern for our carefully crafted categories, individual characters may exhibit asymmetric variation spanning several states, e.g., a base round on one side and straight on the other; this particular type of asymmetry is the only one given a name by the Leaf Architecture Working Group (LAWG 1999), "excavate", shown in Fig. 3.1, though Hickey (1973, 1979) and Dilcher (1974) recommended classifying leaves for both overall and base asymmetry. Nevertheless, these descriptors have proven useful (e.g., for field identification), suggesting that they may represent the types of variation that shape measures should be made to quantify.

Curvature and the Goals of This Study

There are several general approaches to measuring shape that have been applied to leaves: linear measurements, landmark (generally, relative warps) analysis, and outline (usually Fourier or eigenshape) analysis. These methods are all focused on recording a shape and extracting information with some correspondence to common qualitative descriptors. An unanswered—and seemingly unasked—question is: can you accurately analyze leaf shape data without first removing curvature from the data?

By curvature, what is meant here is deviation of the midvein—the axis of bilateral symmetry in leaves—from a straight line. This curvature can manifest as both localized and large-scale variation. Morphometric methods typically control for position, rotation, and size. Curvature and asymmetry represent additional sources of variation, and they may blur the boundaries between character states (e.g., those shown in Fig. 3.1). This can be quite prevalent (Krieger et al. 2007), despite the dearth of qualitative terms, such as "excavate", that directly describe asymmetry. Studies of leaf shape have not addressed leaf curvature, or have simply set a cutoff at which leaves are too curved to include in the analysis; e.g., Paler and Barrington (1995) discarded all leaves where the tip curved more than 2cm from the main axis of the leaf. Approaches which use only the outline (most Fourier and eigenshape analyses) cannot account for curvature; additional information is required, such as the location of the midvein relative to the outline. In the absence of such a reference, asymmetry cannot be discriminated from curvature. Most studies of leaf shape have used measures which do not incorporate the midvein, except at the base and apex (e.g., Whaley and Whaley 1942; Kincaid and Schneider 1983; Young et al. 1995; Premoli 1996; Jensen et al. 2002), and the few that have did not address the issue of curvature (e.g., Jensen 1990; Jensen et al. 1993; Paler and Barrington 1995). West and Noble (1984) developed a method to control for curvature, though not for outline analysis *per se*, drawing a series of lines perpendicular to the midvein in order to properly determine leaf width at all points along the midvein. The limitations of this method are discussed below (see Methods).

The goal of this study was to assess the impact of curvature on the extraction of shape measures from leaf outline data. The method I selected to measure shape is eigenshape analysis. The hypotheses tested were:

1. A sufficiently large sample size, relative to the amount of curvature in the dataset, will allow the eigenanalysis to identify and isolate curvature as a separate source of shape variation.
2. The inclusion of the midvein as a reference curve, appended to the outline data, will allow the eigenanalysis to isolate variation due to curvature.
3. Mathematical removal of curvature is necessary to recover meaningful shape descriptors.

In the first case, when sample size is sufficiently large, the eigenanalysis identifies curvature using the covariance structure of the data. This has not previously been

tested, nor is it clear what sample size is sufficient. Without the midline, it is not possible to distinguish between asymmetry and curvature (see Methods), so H1 seems unlikely. Since curvature is defined by the midvein, H2 seems reasonable: if the midvein is included in the analysis, the eigenanalysis may detect curvature in patterns of covariation between the outline and midvein. However, with no straightforward method to quantify curvature, much less to partition that curvature into orthogonal measures representing different types of curvature-based variation (as an eigenanalysis would be expected to do), it is necessary to mathematically remove curvature from the dataset, to assess the other two approaches.

The ultimate goal of this study is to assist botanists in making informed decisions about how to measure shape in ways that are consistent with classic qualitative terminology. "New methods" for the quantitative description of leaf shape have been shoveled onto the long-suffering botanist for over 30 years (e.g., the early computer program of Dale et al. 1971); I will attempt to review these methods, and describe how some of the newer approaches compare to established morphometric methods. Many morphometric tools are primarily concerned with discrimination among groups. More desirable is the ability to extract meaningful characters; this allows one to ask questions about the location and extent of variation, and to identify the characters which are useful in discrimination of groups. Each method will be assessed with respect to its ability to describe shape using characters interpretable in the context of classic, qualitative descriptors.

Linear Measurements

One approach is to record length, distilling all of the interesting shape information into a single mark along a ruler. The morphospace defined by measuring the length of five leaves is one-dimensional (Fig. 3.2a), extending from zero to infinity. It is not possible to measure leaves with negative length, although you might argue that there are leaves with zero length (their absence identified by leaf scars). In this example, the empirical morphospace (defined by the set of measurements) only extends to 10–20cm or so, depending on the type of leaf. Measuring both length and width creates a two-dimensional morphospace (Fig. 3.2b). A single point in a morphospace is the complete description of shape for an object, for the measurements taken: a point along the axis in Fig. 3.2a describes the length of a leaf; the point (12.75, 7.5) in Fig. 3.2b refers to a leaf that is 12.75cm long and 7.5cm wide. These descriptions become more interesting as you increase the dimensionality of the observations (that is, as more measurements are recorded for each object). Clearly, length and width capture a tiny fraction of leaf shape; compare the leaf silhouettes to the cross-hairs in Fig. 3.2b. In fact, Fig. 3.2b is easy to over interpret: because we have not recorded the point where the two measurements cross or the angle between them (three more measurements), we cannot reconstruct even the shape of the crosses drawn on the plot from the data. The utility of the 2D space with respect to qualitative descriptors is very limited; for example, ovate and obovate leaves (see Fig. 3.1) could occupy the same point in that morphospace. The challenge is to construct a morphospace
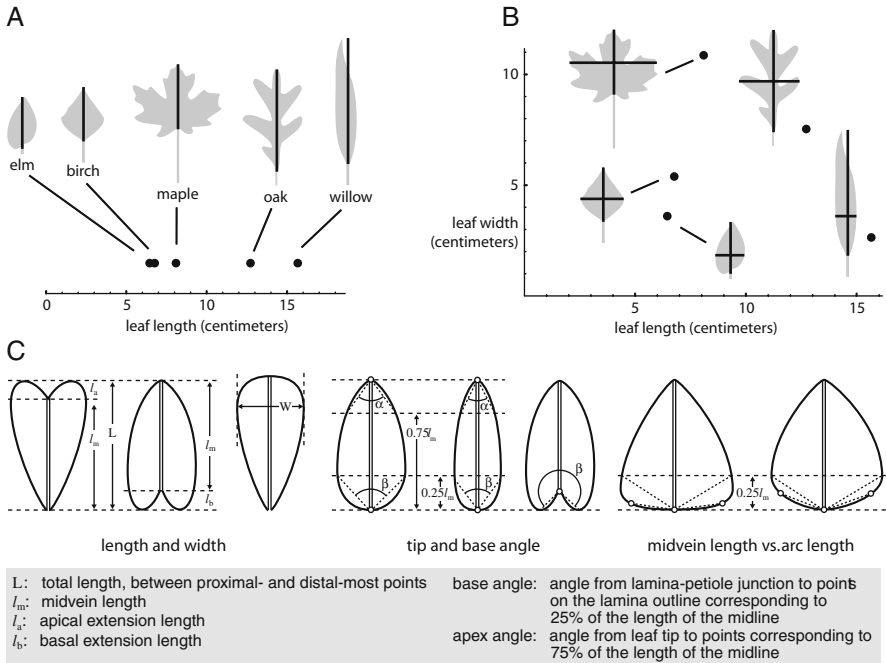
Fig. 3.2 Linear measures of leaf shape. (A) The one-dimensional morphospace arranges leaves in a line based on length of the leaf blade. (B) The two-dimensional morphospace. uses overall length and width at widest point. (C) Linear measurements have become standardized (LAWG 1999). Slight changes to the methodology, such as defining the leaf base using the lowest fourth of arc length along the outline, versus along the midvein, can lead to very different shape measurements. For the two leaves shown at far right, base angle as measured using midvein length is unchanged, whereas using arc length shows the first leaf as having a broader base

with interesting axes, axes encompassing shape variation that tells us something about biology.

Several linear measurements—blade length, width at widest point, petiole length, and their ratios—are commonly used, and these can even be automatically extracted from images (Bylesjo et al. 2008). Rarely, the location of widest part of the leaf is recorded (Dale et al. 1971; Dancik and Barnes 1974). Attempts have been made to standardize these measurements (Hickey 1973, 1979; Dilcher 1974; LAWG 1999; Fig. 3.2c). However, linear measurements tend to be highly correlated (e.g., Sato 1986; Sato and Tsuyuzaki 1988; Usukura et al. 1994), which is a common problem of linear measures and ratios (Zelditch et al. 2004). Even apparently simple measurements such as these are not without their complications. For example, base angle is computed using lines drawn from the base to a normal drawn at one quarter the length of the midline. An alternative, which is arguably more representative of the shapes given in Fig. 3.2c, would be to place the ends of the angle at points corresponding not to a percentage of midvein length, but of outline length. This seems closer to Hickey's (1973, 1979) definition of base—"leaf bound by the lower 25%

of the margin"—and perhaps a better match to an intuitive sense of base angle, but it would be harder to implement in the context of rapid scoring of leaf shape (the point of the LAWG manual). It would also be more sensitive to lobing. However, it could accommodate leaves with l_a or $l_b > 0$ (where the lamina extends past ends of the midvein, e.g., cardioid leaves) without the need for *ad hoc* rules for inconvenient shapes.

One solution has been to elaborate the linear measurements. This has led to "shape" measures such as shape factor ($4\pi \times$ (leaf area)/(leaf perimeter)), feret-diameter ratio ((diameter of circle with same area)/(major axis length), Huff et al. 2003), dissection index (perimeter/($2 \times$ the square root of $\pi \times$ area), Kincaid and Schneider 1983), leaf radian measurements (Meade and Parnell 2003; this is challenging to describe; it's nearly an outline method), and many others. Any combination of linear measurements can be subjected to a principal components analysis (PCA), for example, PCA of length, width at widest point, petiole length, and location of the widest point (Dancik and Barnes 1974). This extracts variance-optimized (the first axis explains the most variance, each subsequent axis explains the same or less), orthogonal measures. While these measures may contain more shape information than simple length or width, they are difficult to relate to useful qualitative descriptors.

Another option is to incorporate additional geometric information, without actually performing geometric morphometric analysis. Linear measurements can be thought of as recording the distances between pairs of points (landmarks) on the leaf. A network of trusses among these points may be used (e.g., Whaley and Whaley 1942; McLellan 1990, 2000), and these tend to achieve better discrimination (McLellan and Dengler 1995) and identify more subtle patterns of variation (Jones 1992; Young et al. 1995) relative to other linear measurement approaches. Euclidean Distance Matrix Analysis (EDMA, Lele and Richtsmeier 1991; Lele, 1993), a truss method that analyzes the complete set of distance measurements among all landmarks, has been shown to introduce structure to the data that interferes with statistical analyses, a significant disadvantage relative to geometric morphometric approaches (Rohlf 2000a, b; summarized in Monteiro et al. 2000). Additionally, linear measures—including trusses—discard much of the positional information present in the points used for measurement (Zelditch et al. 2004).

Investigators will on occasion explicitly state that their linear measures were unable to discriminate taxa which they felt had obvious morphological differences (e.g., Kott and Britton 1982; Sato 1986); this is often implicit in the efforts made to collect linear measures from which no signal could be extracted, presented alongside an image of two taxa which look morphologically distinct. It is in these situations that geometric methods tend to excel.

Relative Warps Analysis

Geometric morphometric methods use the positional information of landmarks, curves, and surfaces. For two closely related species with very similar leaf morphologies, landmark methods may be used for discrimination, where they have been

found to outperform both linear methods coupled with principal components analysis and elliptic Fourier analysis (Jensen et al. 2002). Landmark methods have the potential to isolate characters (Bookstein 2002), though it appears that *a priori* selection of clusters of landmarks corresponding to characters may be preferable (MacLeod 2002b).

For most leaves, landmarks can be placed only at the tip and base (Jensen 1990), though Jensen et al. (Jensen 1990; Jensen et al. 1993, 2002) identified several additional landmarks in closely related species with lobed leaves. However, figures presented in their studies (Fig. 1 in Jensen 1990; Fig. 2 in Jensen et al. 1993; Fig. 1 in Jensen et al. 2002) show a series of leaf forms that cannot be landmarked as easily as their description of methods suggest (Fig. 3.3). The two leaves shown in Fig. 3.3b, c appear very similar in overall shape, but the landmarking technique has enforced a pattern of homology such that the second landmark configuration concentrates most of the variance in shape at a single landmark, the leaf tip. An alternate landmarking approach (Fig. 3.3d) moves the region of expansion to the middle of the leaf, but it still doesn't seem to capture the difference in shape between the two leaves. This may be entirely acceptable if the goal is only to discriminate taxa, without generating meaningful shape measures, or even appropriate, if the lobes were serially homologous, but the expectation is that the whole leaf is homologous (Hageman 1992; Kaplan 1992, 2001).

This issue can become more pronounced: in the dissected-leaf members of *Pleopeltis* s.l.—the fern genus used in this study—pinnae number ranges from 3 to over 160, seemingly precluding the use of landmarks ("pinnae" or "pinnules" is the fern term for "leaflets", units of lamina in dissected leaves). One way around this is to select certain pinna-positions as the relocatable landmark points. Paler
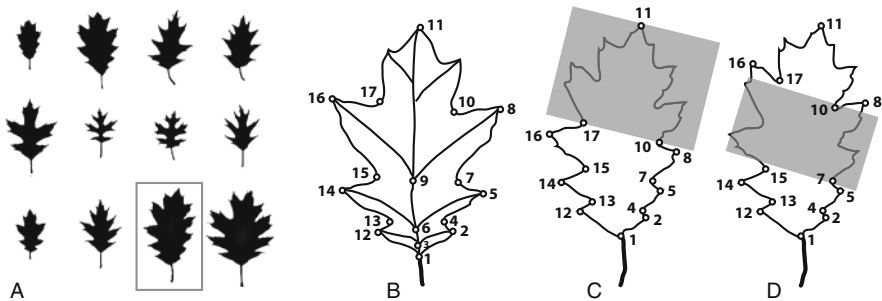
Fig. 3.3 Influence of lobe number on landmark analysis. (**a**) Leaves Jensen et al. (1993) used to illustrate scatterplots, but not landmark location. (**b**) One of two leaves used to illustrate landmark placement (from Fig. 1 in Jensen et al. 1993). The leaves shown in (**a**) exhibit more variable morphology than suggested by this diagram. (**c** and **d**) Possible landmark configurations for one of the scatterplot leaves (gray box in (**a**)), depending on venation pattern. Sizes of the first four lobes (defined by landmarks 1–7 and 12–15) appear comparable. Because of the change in number of lobes, the difference in shape from (**b**) is recorded either as an extension of the tip (landmark 11) or expansion between the first four lobes and the rest of the leaf (illustrated with gray boxes). Neither of these patterns seems to capture the essential differences in shape between the two leaves

and Barrington (1995) used the leaf apex and the tips and bases of a subset of pin-
nae for landmarks: pinna-pairs 1, 2, 3 and 6. This approach was able to identify
four characters to discriminate two fern taxa. However, the degree of homology of
these landmarks will break down with increased variability in number of pinna-
pairs between position 6 and the leaf tip. The power of landmark analysis of simple
leaves is, obviously, limited by discarding the shape information between landmarks
(Jensen 1990; Jensen et al. 2002; MacLeod 1999), which can be addressed instead
through outline analysis.

Elliptic Fourier and EFA-PCA

Their essentially two-dimensional nature means that much of the shape of leaves
can captured in the outline; this excludes venation patterns, but captures both large
and small scale (e.g., marginal teeth) aspects of shape. One widely used quantitative
approach to comparing leaf outlines shapes is elliptic Fourier analysis (EFA; Kuhl
and Giardina 1982), which mathematically decomposes outlines into sinusoidal
functions (Fig. 3.4). Objects are described as closed outlines of x,y coordinates,
and the changes in x and y positions are separately decomposed as periodic func-
tions. Kuhl and Giardina introduced the procedure with outline data recorded using
chain codes, a series of directions that each have one of eight values, corresponding

original outline   1   2   3   4   8   16

number of Fourier harmonics used in outline reconstruction

$$x(t) = a_0 + a_1\cos[t] + b_1\sin[t] + a_2\cos[2t] + b_2\sin[2t] + a_3\cos[3t] + b_3\sin[3t] + a_4\cos[4t] + b_4\sin[4t]$$

$$y(t) = c_0 + c_1\cos[t] + d_1\sin[t] + c_2\cos[2t] + d_2\sin[2t] + c_3\cos[3t] + d_3\sin[3t] + c_4\cos[4t] + d_4\sin[4t]$$

0th harmonic   1st harmonic   2nd harmonic   3rd harmonic   4th harmonic
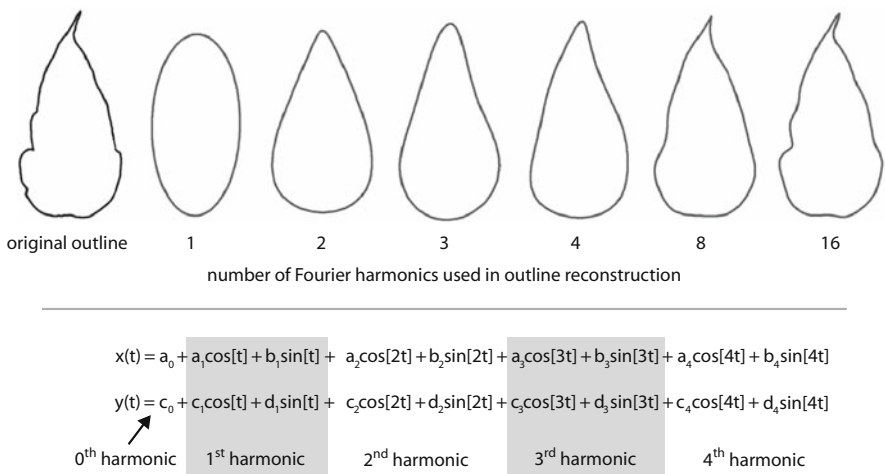
Fig. 3.4 Fourier analysis. The original outline is shown next to reconstructions made with different
numbers of Elliptic Fourier Analysis (EFA) harmonics used. The parametric equations for x and y
are used to generate the reconstructed outlines, based on EFA coefficients a, b, c, and d, which are
calculated from the original outline. The first leaf is reconstructed with the terms for the zeroth and
first harmonics. The fourth leaf uses all the terms shown. As the number of harmonics increases,
the reconstructed outline resembles the original more closely

to 45 degree increments around a circle. However, this is not critical to the method; the chain codes are transformed into changes in x and y, and the raw x,y data can be used instead. Two of the improvements introduced to Fourier analysis by EFA were that points could be unequally spaced, and the outline need not be single-valued (Rohlf and Archie 1984); objects where radii cross the outline in more than one place violate the single-value limitation, which can be an issue for structures with wavy outlines such as lobed leaves.

Shape can be recreated from the resulting Fourier harmonics, with the number of harmonics determining the degree of fidelity to the original form (Fig. 3.4). Each harmonic after the zeroth (the "DC" or "bias" harmonic, corresponding to translation) has four parameters (a_i, b_i, c_i, d_i). These are typically transformed to remove rotation, size, and starting position of the outline. Adjusting the outlines for starting point is reasonable for the original aims of EFA: the automatic recognition of silhouettes. However, in a biological context, it is often possible to specify a relocatable starting point using definitions of homology (e.g., the lamina-petiole intersection), in which case this adjustment will be counterproductive.

For taxonomic discrimination by leaf shape, Fourier harmonics have been found to outperform linear measures of leaf shape (White et al. 1988; Premoli 1996; McLellan and Endler 1998; Rumpunen and Bartish 2002). EFA has been successful for discrimination in hazelnut cultivars (Menesatti et al. 2008), *Betula* (White et al. 1988), and several other crops (Neto et al. 2006). It has been used to demonstrate continuous variation in leaf shape among species in *Sternbergia* (Gage and Wilkin 2008) and *Hoya* (Torres et al. 2008). Ollson et al. (2000) found general agreement between EFA and genetic (RAPDs) matrix distances in Dogrose leaflets. EFA has also been applied to evolutionary studies, quantifying additive genetic variance (Black-Samuelsson et al. 2003) and heterochrony (Pryer and Hearn 2008). One of the earliest uses of Fourier analysis in botany, Kincaid and Schneider (1983), is often incorrectly cited as an EFA study, but they actually use a different Fourier method (Rohlf and Archie 1984).

The goal in using a method like EFA is, typically, to develop a set of meaningful reduced-variance shape descriptors. While Fourier analysis is well-suited to the discrimination of groups, its ability to extract morphological characters to allow interpretation of those differences appears to be severely limited. McLellan and Endler (1998), in their comparison of morphometric methods including fractals and Fourier analysis, describe EFA shape decomposition as "lower order Fourier components correspond to the overall shape, and the higher order harmonics correspond to smaller details…" This should not be misread to suggest that there may be a single high-order harmonic corresponding to, e.g., tip shape. Unlike a linear measurement such as leaf length, which directly captures localized information about developmental processes by indirectly recording the number and size of cells within a region of the leaf, each harmonic in the Fourier decomposition describes a type of diffuse change across the whole shape. Bookstein et al. (1982) argued that the mathematics of Fourier analysis preclude the decomposition of shape into single characters that describe localized change, such as tip and base variation, since localized features are necessarily spread across multiple harmonics. The simplest reconstruction, using a

single harmonic, approximates the leaf as an ellipse. Each subsequent harmonic refines the overall shape, slowly resolving localized characters. Ehrlich et al. (1983) made a compelling case for Fourier coefficients aligning to homologous regions of an outline in a study of foraminifera and suggested that such harmonics should be considered characters. However, studies of leaf form do not seem to support the use of Fourier harmonics in this way. Jensen et al. (2002) felt that the loadings on various EFA components could not be directly related to morphometric characters. Premoli (1996) interpreted only one component, which corresponded to the location of the widest part of the leaf. As the first harmonic to differentiate ovate shape from an ellipse, the second harmonic necessarily describes this type of variation (Fig. 3.4). This seems at best trivially related to ovate-obovate variation in leaf shape.

Even if characters cannot be identified, variance reduction can be achieved by selecting a subset of Fourier harmonics for further analysis. Typically, a qualitative assessment is made of how many harmonics are needed for reconstructed objects to become sufficiently similar to the original shapes. In Fig. 3.4, with eight harmonics, the reconstructed outline begins to look like an actual leaf. However, for specific characters, such as tip or base shape, truncation of the harmonic series eliminates information, and the effects, e.g., on discrimination or disparity metrics, have not been assessed for botanical datasets. Alternatively, the complete set of Fourier harmonics can be decomposed through principal components analysis (EFA-PCA). This approach is not typically used; instead, the PCA is preceded by a reduction in the number of Fourier components (e.g., Pryer and Hearn 2008). It may be that truncation of the Fourier harmonics is appropriate for quantification of leaf features of a periodic nature, such as lobes and teeth, but this has not been studied. An EFA-PCA analysis of the set of dicot leaves from Krieger et al. (2007) using all Fourier components yielded very similar shape models to those seen in the eigenshape analysis from that study (personal observation), suggesting that EFA-PCA might be able to extract shape characters as useful as those seen in eigenshape analysis.

There remain limitations to EFA as currently used. Shape must be described as a closed outline, because a periodic signal is required. However, this is not an inherent limitation for Fourier analysis. Windowed short-timer Fourier transform (STFT) methods applied to ammonoid suture morphology (Allen 2006) look promising for analysis of leaf teeth along open segments of the leaf outline, but it is less clear whether STFT would work for large-scale features in open curves. It is also unclear whether there would be a benefit to segmenting curves using relocatable features (e.g., the leaf tip), as is done in extended eigenshape analysis (see below). This requires separately interpolating different segments of the outline, which is allowable in EFA, because point spacing can be variable. Whether this would improve the results of an analysis, as seen in extended eigenshape analysis, remains to be tested. Despite its origin at the same time as eigenshape analysis, elliptic Fourier analysis has benefited little from developments in the field of geometric morphometrics.

Eigenshape Analysis

Eigenshape analysis was developed by Lohmann (1983) to allow quantification of subtle, continuous variation in shape using curves. It is often referred to as "outline analysis", but the method applies to open or closed curves, which need not be restricted to the edge of an object. Applications include systematics (Ray 1992; Jensen et al. 1993, 2002; Jones 1993; McLellan and Dengler 1995; Premoli 1996; McLellan and Endler 1998; Farley and McNeilly 2000; McLellan 2000; Cannon and Manos 2001; Elewa 2003; Polly 2003; Guralnick 2005), morphological evolution (Schweitzer and Lohmann 1990; Norris et al. 1996; Ubukata 2003; Sille et al. 2004, 2006; Caumul and Polly 2005; Liow 2006), development (Ray 1990), functional analysis (MacLeod and Rose 1993; Sherwood et al. 2005), paleoclimate reconstruction (MacLeod 2002a; Krieger et al. 2007), and species response to environmental variables (Kelly et al. 2001; Teusch and Guralnick 2003; McClain et al. 2004; McClain 2005), to name just a few.

Eigenshape (ES) analysis features four steps: description of shape, alignment of specimens, decomposition into orthogonal functions, and modeling of these shape functions (Fig. 3.5a; Lohmann and Schweitzer 1990). Shape is commonly described using x,y coordinate data for landmarks and curves representing relocatable features (i.e., reproducibly identifiable across all specimens). This step is common to all outline analyses. While seemingly trivial, this is often the most time consuming step, and image resolution may impact the results. Progress is being made both in automated identification of leaves in an image and the extraction of curve data (Belongie and Malik 2000; Belongie et al. 2000, 2002; Wang et al. 2003; Yahiaoui et al. 2006). The degree to which pixelation at the leaf boundary introduces noise to the data may be minimized through smoothing or curve fitting, for example, with spline curves (Chi et al. 2003). In standard eigenshape, the entire curve is interpolated to equally spaced points, either in terms of distance between points or arc length along the curve. These points form the basis for comparison across specimens. Extended eigenshape analysis constrains the curves into segments between landmarks (e.g., leaf tip and base) to enforce homology and allow the eigenanalysis to focus on biologically meaningful patterns of shape variation (Ray 1990, 1992; MacLeod 1999), and it can also accommodate open curves. The landmark constraints on the curve are used to specify regions that are separately interpolated, so all specimens will have equal numbers of points on a given segment. "Standard eigenshape analysis" is really just a subset of extended eigenshape analysis, and "eigenshape analysis" should now be taken to be the latter.

Specimen alignment is a procedure that removes size, position and rotation from curve data. This is typically achieved in eigenshape analysis by conversion to phi functions (Fig. 3.5b; Zahn and Roskies 1972). Phi functions describe curves as functions of cumulative angular change (Fig. 3.5b). The coordinate data are transformed to changes in angle required to follow the curve; this information is unchanged under translation, rotation, and scaling. Reconstructing the original shapes requires both the phi functions and the segment lengths (or, at least the relative lengths
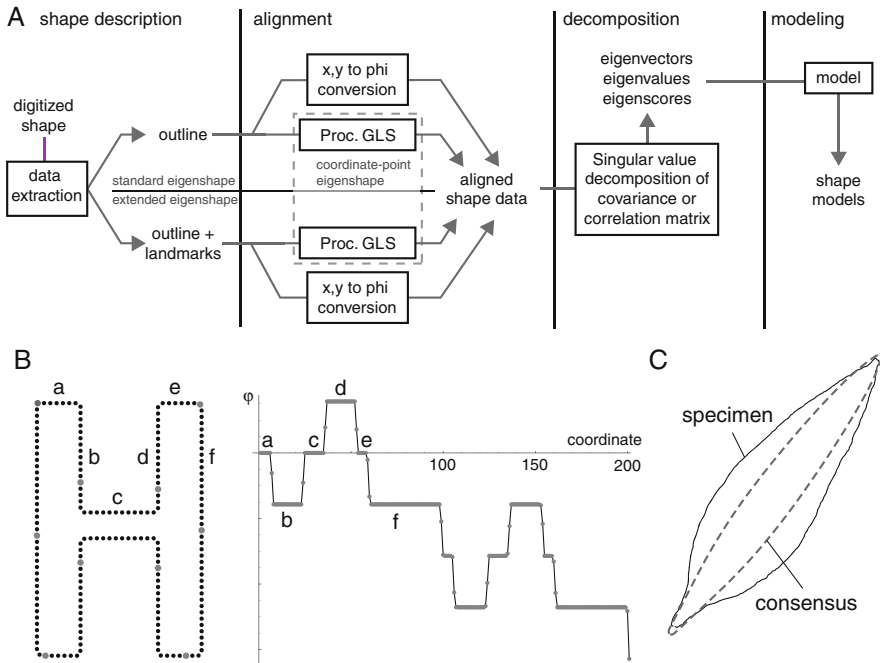
Fig. 3.5 Eigenshape analysis, phi functions, and Procrustes GLS superposition. (**a**) Overview of eigenshape analysis. Shape description transforms an image (or, an actual object) into analyzable data, e.g., outline x,y coordinates. Alignment removes position, rotation, and scale from the objects, leaving shape information. The eigenanalysis occurs in the decomposition stage, generating a set of variance-optimized, orthogonal measurements. Modeling, which is possible when complete object geometries are analyzed, allows the visualization of the morphospace. The four steps could equally be applied to relative warps analysis. (**b**) Phi functions describe the angular change between tangents to points along a curve, expressed as a cumulative function. A sample closed outline is shown, starting at segment "a". Each flat region on the phi function, a plot of coordinate number versus phi, corresponds to one section of the "H" (see labels a-f). The phi function describes cumulative change, so it remains constant when the angle between individual line segments does not change. Phi functions for closed, clockwise curves begin at 0.0 and end at -2π. Redrawn (and recomputed) from Fig. 3.2 of Zahn and Roskies (1972). (**c**) Procrustes generalized least squares (GLS) superposition. A single leaf is shown overlaid on the sample mean shape. Specimens are rescaled to unit centroid size, then rotated and rescaled to minimize the distances between object outline points and the corresponding points on the mean shape (which is recomputed after each round of reorientation)

of segments, if they are size-standardized). The latter information is not typically included in the eigenanalysis. If the outlines were interpolate to equidistant points, the x,y data may be reconstructed from the phi functions; this may not be the case if interpolation is carried out using equal distances along the curves, though the latter approach is truer to the Zahn and Roskies formulation of phi functions.

In the decomposition step, a correlation or covariance matrix of these phi functions is subjected to a singular value decomposition, and the resulting eigenvectors

describe a morphospace. The eigenvectors (typically called "eigenshapes" in eigenshape analysis) represent a new set of variance optimized, uncorrelated shape measures.

The final step is modeling, visualizing shape for points in the morphospace. This is a particularly useful feature of eigenshape analysis, allowing you to identify the nature of variation along individual axes. Models can, in fact, be created for any space that is built through the linear combination of variables (Rohlf 1996), such as the eigenanalysis of EFA-PCA, but they are rarely used outside of relative warps and eigenshape analysis. Shape variation along an eigenshape axis (hence, the variation represented by that eigenshape) can be visualized by varying the amplitude of the eigenshape for that axis and adding this to the consensus shape (Lohmann 1983; MacLeod 1999). The consensus can be defined in various ways, typically as the sample mean shape computed during alignment. These models are used to determine the nature of shape variation along individual axes in the shape space, and interpretations of the modeled variation can be confirmed by examining the distribution of object shapes (or their qualitative descriptions) along individual eigenshape axes (e.g., the results shown in Fig. 3.9).

Coordinate-point eigenshape (CPES), which is only beginning to be adopted, dispenses with the phi conversion and instead compares the coordinate data of specimens, in two or three dimensions (MacLeod 1995, 1999, 2001). Like phi function based eigenshape analysis (phi-ES), curves may be constrained with landmarks. Unlike phi-ES, alignment is achieved through Procrustes Generalized Least Squares (GLS) superposition (Rohlf and Slice 1990), making this technique mathematically equivalent to an unweighted relative warps analysis (e.g., Rohlf 1993). CPES can also include isolated landmarks along with curve data. Clearly, there is no way to describe an isolated point in a framework of cumulative angular change. It is trivial to extend the technique to three dimensions by simply adding another dimension to the calculations (for landmarks and curves; it is more complicated for surfaces; see MacLeod 2008); the same can be done for phi-ES, with a bit more effort (MacLeod 1999). The extension to 3D surfaces is termed eigensurface analysis, MacLeod 2008; Polly and MacLeod 2006.

Sampson et al. (1996) explored a morphometric method for ventricular outlines very similar to coordinate-point eigenshape analysis. Interpolated outlines were aligned using Procrustes superimposition, but then the points were allowed to slide along the outline before another round of Procrustes alignment. This approach was later refined by Bookstein (1997) as sliding semilandmarks, and it suffered from predating the development of extended eigenshape analysis (see the discussion of their study in MacLeod 1999). To the degree the Sampson et al. (1996) results seem useful, it would be better to explore Bookstein's method, or use CPES. It is an open question, the degree to which sliding semilandmarks distort the raw coordinate data.

Eigenshape analysis, as applied to leaf outlines, has been shown to outperform Fourier analysis for discrimination of taxa (McLellan and Endler 1998). Shape variation can be localized to regions as small as a pair of curve segments or coordinate points, and this size limit is set by the resolution of the outlines. Therefore, eigenshape analysis has the capacity to extract very small, localized shape characters. It

also has the potential to isolate leaf shape characters (e.g., tip shape) with minimal *a priori* selection of what to measure (Krieger et al. 2007).

Other Eigenanalysis-Based Methods

Several recently developed methods for leaf analysis are confusingly similar to existing geometric morphometric methods. All of these methods share eigenanalysis, which means they can utilize powerful shape visualization tools, but they typically fall short in their approaches to shape description. These methods are described here in the hopes of limiting further development of *ad hoc* methods, and referencing the ever-growing suite of approaches to the geometric morphometrics framework.

Langlade et al. (2005) developed a method to study allometry in *Antirrhinum* leaves. A set of 19 points were placed on each leaf. Four were relocatable: the base and tip, and two points at the distal limit of the pedicel (where the lamina constricts to the midvein). These landmarks describe the positions of homologous regions of the leaf; the same is not true of the next two landmarks, placed on the outline at the widest point of the leaf. The remaining semilandmarks were interpolated along the lamina (four on each side) and midvein (five). It would have been preferable to interpolate the leaf outline using the first four relocatable landmarks to segment it into two curves along the lamina and a third along the midvein. Additionally, describing the lamina with so few points is unnecessary, because variance reduction is better achieved in the principal components analysis to follow, and this downsampling is likely to miss subtle patterns of shape variation. Leaves were rotated and translated to make the midvein horizontal and place the centroid at the origin. The alignment is not a Procrustes superimposition, nor is it exactly a use of Bookstein coordinates (Bookstein 1986), which would rotate, scale, and translate to place the ends of the baseline at (–0.5, 0) and (0.5, 0). They successfully use the resulting shape space to address interesting questions, but the danger in using such an *ad hoc* method is that you don't know whether the method used to construct the morphospace has introduced structure to your data (Rohlf 2000a). This begs the question of why they chose to ignore well-established methods.

The LeafAnalyser program (Weight et al. 2008) interpolates leaves to a user-specified number of equally-spaced points (it's not clear if they are equidistant or have equal lengths of the original outline between them) and translates them to have the centroid at the origin and rotating the outline so the leaf tip is on the positive y-axis (above the centroid). They describe an analysis of *Antirrhinum* leaf data standardized in this way. The alignment is not a Procrustes superimposition. Using the tip and centroid as a baseline means that it is closer to using Bookstein coordinates than the Langlade methods, but the baseline may suffer from instability due to the centroid, the position of which can vary depending on the interpolation of the original data. Additionally, aligning the leaves in this way transfers variance away from the baseline, hence the leaf tip, to other semilandmarks. As the number of points is decreased, this will decrease the ability of the eigenanalysis to recognize

variation in tip shape. The aligned leaves were then mean-centered, but not rescaled, and subjected to a principle components analysis. They describe the first principal component as size, despite the allometry clearly visible in the shape models for that axis. The initial automated outline extraction provided by LeafAnalyser is very useful, but the analysis should instead use a Procrustes superimposition and, if scale is to be left in, specimens should be scaled to the square root of centroid size. As with Langlade et al. (2005), it is unclear why they ignored existing geometric methods.

The Langlade et al. (2005) approach was updated by Bensmihen et al. (2008) in a study of *Arabidopsis* and *Antirrhinum* leaf mutants. The leaf outlines were interpolated following an (unreferenced) extended eigenshape protocol, interpolating among the first four landmarks described above (tip, base, and two points at the limit of the pedicel). Leaves were then aligned using Procrustes superposition. Therefore, they applied coordinate-point eigenshape analysis, without actually recognizing it as such.

Other Descriptions of Leaf Form

Architectural approaches specify parameters such as frequency and angle of branching in stem or vein networks. These approaches are focused primarily on theory, generating shapes with apparent similarity to real plants (Niklas 1978, 1994a, 1999; Stein and Boyer 2006), though Niklas has certainly approached model development critically (e.g., Niklas 1994b). Simulation for the sake of making leaf-like objects is taken farther still with L-systems (Prusinkiewicz and Lindenmayer 1990; Hammel et al. 1992; Prusinkiewicz et al. 1996) and other artificial modeling approaches (e.g., Mundermann et al. 2003). Recent developments in this area have potential to illuminate the origins of leaves (Stein and Boyer 2006), but the statistical behavior of the resulting morphospaces is poorly known. The venation of simple leaves can be directly compared to the branching pattern of dissected leaves (Plotze et al. 2005). This type of approach would be problematic for most herbarium specimens, where veins are not typically visible and the leaves cannot be cleared, eliminating a significant source of data. There is also the open question of how venation relates to outline shape. It would be useful to quantify the patterns of covariance between these two features of leaf form; until this has been accomplished, leaf outline analysis can't be dispensed with.

The classic approach of marking an immature leaf surface with a grid of points (e.g., Avery 1933) or more recent marking of mitotic figures (Zurakowski and Gifford 1988) are well-suited to describing the location and degree of growth, but don't describe shape *per se*.

Fractal dimension is a single metric which has been used to quantify margin roughness (McLellan and Endler 1988) and branching patterns (Campbell 1996). The fractal description of margin roughness is highly correlated with dissection index (the standardized ratio of perimeter to square root of area), which is much easier to calculate (McLellan and Endler 1988). Neither measure is describing localized morphometric characters. Multiscale fractal dimension has been shown to generate

better discrimination than Fourier analysis and linear measures (Plotze et al. 2005), but discrimination is not the same as decomposition of shape for character extraction. Fractals do not describe shape in the same way as a qualitative descriptor or outline-based measure; rather, they look at patterns of self-similarity and generate a compact description of form (Campbell 1996), either as a single value of dimension or as a curve, in the case of the multiscale fractal dimension. In general, single-parameter measurements of shape are problematic. It is unlikely that shape can be distilled to a single, meaningful value, and it is often the case that similar values result from very different shapes. Like architectural approaches, fractals have been used primarily to generate models of fern leaves (Campbell 1996), and the rare efforts to tie these models to real shapes rely more on subjective comparison than empirical analysis of shape variation (e.g., Heggie and Zodrow 1994). Neither architectural nor fractal methods seem appropriate to the decomposition of leaf shape into interpretable characters.

Materials and Methods

This study focuses on simple-leaves species in the fern genus *Pleopeltis* Humb. & Bonpl. ex Willd. (Polypodiaceae). Leaves are the primary source of macroscopic morphological diversity in ferns. Over the course of evolutionary history, ferns have spanned the morphological variation represented by all vascular plants, and most of that variation is still present in extant lineages (Boyce and Knoll 2002). Most fern species have pinnate-determinate leaves—the archetypal fern frond—but simple-leaved species are distributed throughout the leptosporangiate ferns. Approximately 10% of fern species have simple leaves, and approximately 17% of genera contain simple-leaved species (Tryon 1964). Among the basal fern lineages, species with simple leaves are found in the filmy ferns (8% of species) and Schizaeaceae (12%), and they are absent from the Osmundaceae and Gleicheniaceae (Tryon 1964; Iwatsuki 1990; Kramer 1990b–e). Within the core leptosporangiate ferns, simple leaves are found in the water ferns (21%), Cyatheaceae (<1%), and in approximately 10% of species in the hyper-diverse polypod clade (Tryon 1964; Kramer 1990a, c; Schneller 1990a, b). Morphology-based taxonomies have traditionally recognized groups of ferns sharing large-scale leaf forms: simple, pinnatifid (a term for lobed leaves in ferns), pinnate, and more highly-dissected pinnate leaves. As the phylogenetic relationships among ferns have been reexamined using molecular data, our assessment of the evolutionary lability of leaf form has changed dramatically. Molecular systematics has recently provided evidence for rampant homoplasy in leaf form, spanning the range of simple, pinnatisect, pinnatifid, and pinnate morphologies (Murakami et al. 1999; Ranker et al. 2004). Thus, it can be particularly useful to be able to make precise quantitative measurement of simple leaf forms.

Within the polypods, the genus *Pleopeltis* Humb. & Bonpl. ex Willd. (Polypodiaceae) has until recently been circumscribed as a small group, with 30–40 species (Mickel and Beitel 1988). The distribution, shape, and coloration of scales

have traditionally been used to distinguish both species and varieties in the "scaly polypods" (*Pleopeltis* s.l.). Recent molecular work by Haufler (C.H. Haufler, 2003, University of Kansas, personal communication) and Schneider (Schneider et al. 2004 ; H. Schneider, 2006, personal communication) and preliminary monographic work by Smith (A. R. Smith, 2004, University of California, Berkeley, personal communication) has recircumscribed the group to include some members usually treated in *Polypodium* L., *Neurodium* Fée, and *Dicranoglossum* J.Sm. Taxa were selected to represent most of the widespread species from Mexico to northern South America. Lists of the species used in this study are given in Table 3.1. I included all eight simple-leaved species of *Pleopeltis* used by Hooper (1994), though I have not split *Pl. astrolepis* into 2x and 4x groups, as this information was present on only a few of the specimens used in that study. Overall, sample sizes are comparable to or slightly larger than other published eigenshape analyses. Taxa with several varieties were included (*Pl. polylepis* and *Pl. polypodioides*). Two taxa which may

Table 3.1 Fern species used in this study. The number of individuals is the number of herbarium sheets, each a separate accession, from which one or more leaves were sampled. Additional information is given in Krieger (2007). Pl. = *Pleopeltis*, Po. = *Polypodium*, M. = *Microgramma*, N. = *Neurodium*. *N. lanceolatum* is likely to be recircumscribed into *Pleopeltis*; *M. percussa* (previously included in *Pleopeltis*) and *Pl. fuscopunctata* are likely to belong to *Microgramma*

Species	Leaves	Individuals
Pl. astrolepis (Liebm.) E. Fourn., Mexic. Pl. l:87. 1872.	64	28
Pl. complanata (Weath.) E.A. Hooper, Amer. Fern J. 85:76. 1995.	56	23
Pl. conzattii (Weath.) R.M. Tryon & A.F. Tryon, Rhodora 84: 129. 1982.	47	19
Pl. crassinervata (Fée) T. Moore, Index Fil. 345. 1862.	65	22
Pl. fructuosa (Maxon & Weath ex Weath.) Lellinger, Proc. Biol. Soc. Wash. 89: 722. 1977.	100	24
Pl. fuscopunctata (Hook.) R.M. Tryon &A.F. Tryon, Rhodora 84: 129. 1982.	18	5
Pl. intermedia M. Kessler & A.R. Smith, Candollea 60: 275. 2005.	86	25
N. lanceolatum (L.) Fée, Hist. Vittar. Pleurogr. (Mém. Foug. 3): 28. 1852.	45	20
Pl. macrocarpa (Bory ex Willd.) Kaulf. Jahrb. Pharm. 21:41. 1820.	66	29
Pl. mexicana (Fée) Mickel & Beitel, Amer. Fern J. 77:21. 1987.	65	29
Pl. panamensis (Weath.) Pic. Serm., Webbia 23: 189. 1968.	21	6
M. percussa (Cav.) de la Sota, Physis, Secc. C. 44(106): 28. 1986.	51	33
Pl. polylepis (A. Roem. ex Kunze) T. Moore, Ind. Fil. 348. 1862.	58	20
- var *erythrolepis* (Weath.) T. Wendt, Amer. Fern J. 70:9. 1980.	49	18
- var *interjecta* (Weath.) E.A. Hooper, Amer. Fern J. 85:79. 1995.	61	17
Pl. stolzei A.R. Smith, Candollea 60:282. 2005.	11	5
Pl. wlesbaurii (Sodiro) Lellinger, Proc. Biol. Soc. Wash. 89:723. 1977.	74	16
	Total: 937	339

not belong to *Pleopeltis* s.l. were included: *Microgramma percussa*, which has pre-
viously been included in *Pleopeltis*, and *Pl. fuscopunctata*, which will likely be
placed in *Microgramma* (A.R. Smith, 2004, personal communication).

   Pleopeltis is an evolutionarily young group, estimated as having radiated less
than 10 million years ago (H. Schneider, 2004, personal communication), lessening
the chance that extinction of lineages will result in overestimating the morphological
disparity that exists between extant, putative sister taxa. The molecular phylogenies
support at least two origins of simple leaves within clades of dissected leaves. The
recent diversification of *Pleopeltis* s.l., coupled with the apparent multiple origins
of simple leaves, make this an excellent group for the study of the evolution of leaf
shape.

   The specimens used in this study were digitally photographed from a total of
1,432 herbarium sheets (a number which includes dissected leaf specimens, used in
a combined study; Krieger 2007) from the Missouri Botanic Garden (MOBOT),
the University of Colorado Herbarium (COLO), the New York Botanic Garden
(NYBG), the Field Museum, the Jepson Herbarium at UC-Berkeley (JEPS), and
the Herbario de la Universidad de Guadalajara Instituto de Botánica (IBUG). COLO
images include personal collections made in Costa Rica over the course of this study.
Images of an additional 22 herbarium sheets were downloaded from the Natural
History Museum, London, FMNH, MOBOT, and NYBG. For species photographed
at FMNH, JEPS, MOBOT, and NYBG (see Krieger 2007, for a complete list), all
sheets were examined for each species, and all sheets with material that appeared
usable were photographed. Of the herbarium sheets that had usable leaves, typi-
cally half or more of the leaves were damaged or obscured, and roughly half of the
leaves that remained were highly curved. Therefore, it was critical to find a means
to include the curved leaves.

Image Processing

Individual leaves and scale bars were digitally isolated from herbarium sheet images
and converted to silhouettes in Photoshop (various versions, Adobe Systems, Inc.)
for further processing (Fig. 3.6). The lamina was made 50% gray, the midvein
and petiole black, with a gap in the petiole basal to the lamina-petiole junc-
tion. This allowed manipulation of settings in tpsDig (ver 1.40, F.J. Rohlf, 2004;
http://life.bio.sunysb.edu/morph/) to alternately capture the midvein and lamina out-
lines. Three tpsDig format data sets were created: landmarks, lamina outline, and
midvein outline (Fig. 3.6c). The landmarks file included one tip landmark and four
petiole landmarks. Two of the petiole landmarks were relocatable across speci-
mens (a better description than "homologous", when describing point locations,
even if they represent the junction between homologous structures): the lamina-
petiole junction and the petiole-rhizome junction. The other two were placed to
best approximate petiole length (for a separate study). In a small (<1%) fraction of
leaves, high petiole curvature resulted in an underestimation of length, but a larger
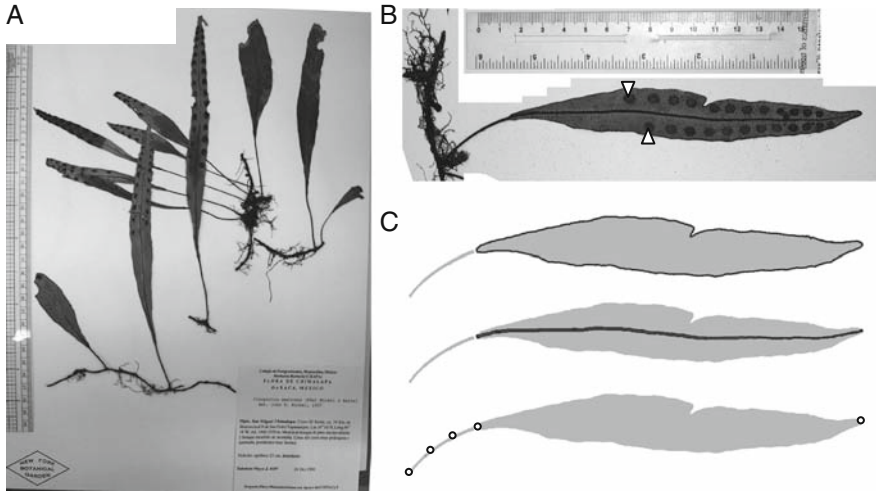
Fig. 3.6 Digitization of leaves. (**a**) Original herbarium sheet with simple leaves. Leaves were typically overlapping and often damaged. (**b**) Isolated simple leaf from another sheet, with scale bar. (**c**) Data extraction from silhouetted leaf. Three types of data were saved for each simple leaf: lamina outline, midvein outline, and landmarks. Four landmarks were placed along the petiole to measure length, including petiole-rhizome junction and petiole-lamina junction. A fifth landmark was placed at the leaf tip. Additional landmarks could be placed in the centers of the most basiscopic sori (arrows in (**b**))

number of petiole landmarks could be used if greater accuracy is desired. The lamina and midvein files were digitized as closed outlines of at most 6,000 points.

Eigenshape Analysis

All analyses were performed in Mathematica (version 7.0, Wolfram Research Inc.); the straightening and phi-based eigenshape code may be downloaded from the morpho-tools website (http://www.morpho-tools.net). The three leaf files were imported and visualized. When initially digitized the ends of the midvein and lamina outlines were not necessarily at the tip or base (petiole-lamina junction) landmarks; the analysis first identified the sections of the outlines corresponding to the landmarks, then reordered the points to place the outline and midvein ends at the base landmark (this allows for the input of messy data, making the method easier to apply). The outlines were split into two segments each for the variance reduction associated with extended eigenshape analysis of homologous segments. To ensure that just the lamina was being compared, lamina outlines were pruned at the intersection with the midvein. This was necessary as several of the species used in this analysis had asymmetric extension of the lamina at the tip and base.

Though performed in Mathematica, the eigenshape analyses yielded results identical to MacLeod's FORTRAN programs, within rounding error (see Krieger 2007

for a comparison). Both the phi and coordinate-point analyses used mean-centering, where the consensus shape was subtracted from all specimens. This centered the analyses on the consensus; normally, the first eigenshape represents an axis of similarity (Lohmann 1983), explaining a large percentage of the variance, often greater than 90%. Mean-centering eliminated that axis, making the results more comparable to a relative warps analysis (identical, for CPES). For coordinate-point eigenshape, the mean centered specimens were aligned using Procrustes GLS superposition. Otherwise, both analyses were identical, as illustrated in Fig. 3.5a.

Analysis 1: Sample Size

To test the effects of sample size, 500 runs were performed on random subsamples, from ten specimens up to the total sample size. Eigenshape analysis was performed on the phi functions or aligned coordinate-point data to extract specimen scores for the subsampled datasets. Alignments varied with each sample in the CPES analyses. Mean centering was performed as in the complete analysis, using a mean shape computed from each subsample. To assess how quickly the analyses converged on the final morphospace, a correlation was computed between scores of objects in the subsampled space and their scores in the full analysis. The absolute value of the correlation was used, because the sign of individual eigenshape axes is irrelevant (being mutually orthogonal, individual axes can be reversed with no effect on the analysis). Resampling was performed for the first five eigenshape axes. ES30 was used for comparison, because this axis explained a small amount of the variance and would be expected to be highly sample dependent, thus more variable for small sample sizes. In principal, the eigenshapes could be compared between the original analysis and subsamples, but it is more appropriate to compare specimen scores, which represent measurements on the actual objects. Several axes explained very similar amounts of variance (e.g., ES3 and 4; Fig. 3.9), so correlations were computed with the two neighboring axes, and the maximum value was used. Failing to do this would greatly underestimate the similarity between the subsampled and full morphospaces.

Analysis 2: Appending the Midvein

A second eigenanalysis (leaf analysis 2; LA2) was performed using a combination of raw outline and smoothed midvein data. To compare analyses before (LA1) and after (LA2) addition of the midveins, shape models were generated for the first seven eigenshape axes. Each of the eigenshapes represents a type of shape measurement, and the score of a specimen on that axis is the value of the measurement. Thus, correlations of scores (Pearson's product-moment correlation coefficient) between the two analyses serve as estimates of how similar the measures are. For each axis in LA2, correlations of scores for the first seven axes in LA2 with scores along the first

25 axes in LA1 were computed. As above, the absolute value of the correlation was used. The highest value was noted, and values less than 0.5 were dropped as relatively weak correlations (a fairly strict cutoff, but similarity seemed to drop quickly below a correlation of 0.7).

Analysis 3: Straightening

To assess the benefits of removing curvature, leaves were mathematically straightened. A number of options were examined for straightening the lamina outline. Straightening presents several geometric issues, including reducing error due to small-scale noise in the reference curve, managing point densities, preserving point order, and retaining features such as invaginations and cardioid bases and tips. In this study, the reference curve used for straightening was the midvein, which was present, easily visible, and extended all the way to the tip for all leaves. To reduce the effect of noise due to pixelation and error during tracing, the midveins were smoothed by a ten-point average. The midveins averaged 1,000 points long, so the ten-point average retained large-scale curvature and much of the small-scale curvature, as well. A five-point average did not adequately remove small-scale noise. West and Noble (1984) also performed midvein smoothing before extracting width measures, using a five-point linear regression.

The straightening approach used by West and Noble (1984) to collect linear measures is the most intuitive automated approach: at each point along the smoothed midvein, a line is drawn perpendicular to the tangent at that point. The width at each midvein point could then be used to reconstruct the leaf in a straightened form (though West and Noble did not perform this step). The normal lines often cross one another (Fig. 3.1 in West and Noble 1984), so the mapping between midvein and outline is imperfect; the algorithm favors information from the midvein over that of the outline. Effectively, the order of outline points is ignored. The algorithm also truncates cardioid bases and tips, and it follows only half of the contour of an invagination. The West algorithm could be modified to retain all of the points where the normals cross the leaf margin, but this would amplify the issues with point ordering. The sample analyzed here included a small number of leaves with cardioid (emarginate) tips and many leaves with deep furrows along the margin, a potentially taxonomically informative feature that needed to remain in the analysis.

The algorithm used in this study gave priority to the lamina outline, to ensure that all points along this curve were present in the straightened outline and retained in the correct order. The lamina was straightened using a series of vectors referenced to tangents along the smoothed midvein (Fig. 3.7).

For each point along the lamina, a vector was calculated from the closest point on the midvein. The tangent to the midvein at this point was approximated as the vector between the midvein points before and after the current point; because of the ten-point smoothing small-scale noise had a minimal effect on the tangent vectors. The angle between this tangent and the x-axis was then used to rotate the midvein-lamina vector (since all points along the straightened midvein will have the x-axis as
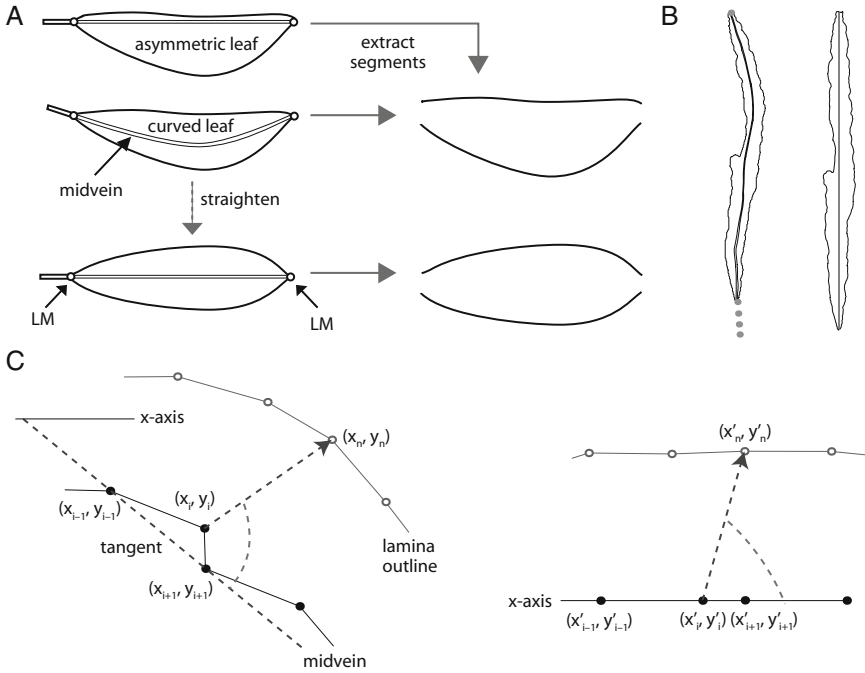
Fig. 3.7 Curvature and straightening. (**a**) Curve segments extracted from unstraightened leaves are misleading. The unstraightened segments portray a leaf that is flat on one side and curved on the other. This suggests an asymmetric leaf, when the actual shape is better captured after straightening using the midvein as a reference. (**b**) A sample leaf before and after straightening. (**c**) Each point along the lamina outline (open circles) is referenced to its closest point on the midvein, as a midvein-lamina vector (from x_i, y_i to x_n, y_n). The tangent to the midvein at that point is approximated as the line between the bounding points (x_{i-1}, y_{i-1} and x_{i+1}, y_{i+1}). Since this is the smoothed midvein, these two points are actually the average of the surrounding 12 points, so the variation in segment lengths here is an exaggeration. (**c**) The midvein points are moved to the x-axis. The midvein-lamina vector is translated to the corresponding new midvein point along the x-axis. The angle between this vector and the tangent is used to rotate the translated vector, giving the straightened lamina outline point

their tangent). The midvein points were placed along the x-axis, spaced according to the Euclidean distances between points along the midvein. The rotated vectors were used to generate the outline points relative to the transformed midvein points. This process was performed for all lamina outline points for all leaves in Mathematica ("Leaf curve Step 2 v1.7.nb", available on morpho-tools.net).

The midvein was not used in leaf analysis 3 (LA3), other than as a reference for straightening. After the straightening procedure, the midveins varied only in length, which, after resizing to unit centroid size, was entirely a function of leaf outline shape, and the resulting base-tip distance. There are no apparent benefits to retaining this redundant information in the analysis.

When a curve is straightened, the new straight region has a very high density of points. For a small number of leaves (roughly 30 of 937), this process

generated small loops in the outline. While seemingly inconsequential, they significantly impacted the amplitudes of the phi functions. Phi functions describe cumulative angular change; therefore, a closed outline is represented by a phi function that starts at 0.0 and ends at 2π or -2π (depending on whether it is counterclockwise or clockwise). A clockwise loop added to a clockwise closed outline results in a phi function that goes from 0.0 to -4π; a counterclockwise loop on the same outline results in a phi function from 0.0 to 0.0. These differences become the primary source of variance in the data, so they must be removed before the eigenanalysis. To identify these loops, phi functions were examined for all straightened leaves. A first pass identified all functions that didn't end at 2π. Loops were then localized by looking for large $(+/-2\pi)$ deflections to the phi function and digitally removed by dropping the offending points (essentially, clipping off the loops). After this procedure, a preliminary eigenanalysis revealed the presence of additional loops: the first eigenshape showed a pattern of variation where the two halves of the leaf unfolded from the tip, to a degree that the ends rotated 180 degrees and met again, completely inverting the form. This clearly represented a mathematical problem and not natural variation in geometry. A careful check revealed several outlines with loops in both directions on the same outline, thus obscuring their effects. The outlines were then re-interpolated to ensure equal spacing of outline points. This process did not otherwise change the shape of the leaves. The "de-looped" outlines were also used for coordinate-point eigenshape, though this procedure was not necessary, because loops have very little effect in coordinate-point eigenshape analysis. This illustrates a key difference between CPES and phi-ES: a loop—as a 360 degree angular change—represents a major shape feature in a decomposition of phi functions. For an analysis of coordinate locations, a loop is not very different from a bump (both are just clusters of points in space), the latter feature having minimal impact in either type of analysis.

A related approach to straightening was also explored. Instead of a vector between a lamina point and its closest midvein point, a vector was drawn from each successive midvein point to each lamina point in turn (that is, "i" vectors, from midvein point "i" to lamina point "i"). As with the preceding method, this vector was rotated using the angle between the tangent and the x-axis. While seemingly very similar to the "closest point" method, this method returned many more artifacts. The compression of highly curved sections of the outline resulted in many loops and invaginations. These artifacts would become more pronounced with midveins less well-sampled than the specimens in this study. In terms of leaf growth, it seems preferable to associate each outline point with the closest midvein point, instead of adopting the arbitrary requirement of using all points in the midvein.

To test the degree to which removing curvature changes the variance structure of a morphospace and improves interpretability of eigenshape axes, absolute values of correlations of scores (Pearson's correlation coefficient) among the three analyses (between LA1 and LA3, and LA2 and LA3) were computed. As above, for the first seven eigenshape axes in LA3, absolute values of correlations of scores in LA3 with scores along the first 25 axes in LA1/LA2 were computed. The highest value was noted, and values less than 0.5 were dropped.

Results

Leaf analysis 1: sample size

Results for the randomization trials are shown in Fig. 3.8. As sample size increased, for both phi and coordinate-point eigenshape analysis, scores of specimens on the first five eigenshapes quickly converged on those from the full sample. ES1-5 reached a mean correlation of 0.8 by a sample size of roughly 200 for both analyses; by contrast, at a sample size of 800, ES30 (which is expected to be highly sample dependent) had a mean correlation of 0.4 in the phi analysis and just over 0.8 in the CP analysis. Therefore, the patterns seen in ES1-5 are not simply artifacts of the resampling procedure.

Phi-ES analyses were slower to converge, but for the first two eigenshapes the lower bound of the 95% confidence interval for the correlation coefficient quickly exceeded 0.9. Sample dependence increased with subsequent axes, as the variation described became increasingly localized (see the shape models, below). This is the case because, ultimately, each specimen has to be described by scores on the eigenshapes, but increased sample size decreases the impact of single specimens on early eigenshapes. For the first five eigenshapes, the mean correlation coefficient was greater than 0.9 by a sample size of 400 (versus 937 total), with the lower bound of the 95% C.I. roughly 0.8 or more. By contrast, for ES30, the correlation was very low until a sample size of 800. This suggests that whatever benefits might be imparted by high sample size should be seen in this analysis.

The results further suggest that C-P eigenshape analysis of leaves should have a minimum sample size of 200, corresponding to a mean correlation coefficient of 0.9 or greater, through ES5. Phi eigenshape analyses appear to require 50% more specimens for similar results. These values will depend entirely on the degree of similarity across specimens in the sample and patterns of shape variation in the sample.

Leaf Analysis 1 vs. 2: Appending the Midvein

Comparisons of shape models between LA1 (raw outline data) and LA2 (raw outline data plus smoothed midveins) are shown in Figs. 3.9 and 3.10. The first seven eigenshapes were examined. There is no set rule for how many axes to consider; for this dataset, ES7 represented less than 2% of the variance explained in all four of the analyses (LA1 and 2 for phi and CPES), but it was still relatable to qualitative shape descriptors (discussed below).

For the analyses of phi data, there was very little change in the first seven eigenshapes after the addition of the midvein. This was evident from inspection of the models (Fig. 3.9), and the high Pearson's correlation coefficients (PCC) provided additional support. All of the axes in LA2 were represented in LA1: ES1-5 had PCC's greater than 0.9, and ES6 was only slightly lower at 0.839. The variance explained for individual shape measures changed between analyses, thus the
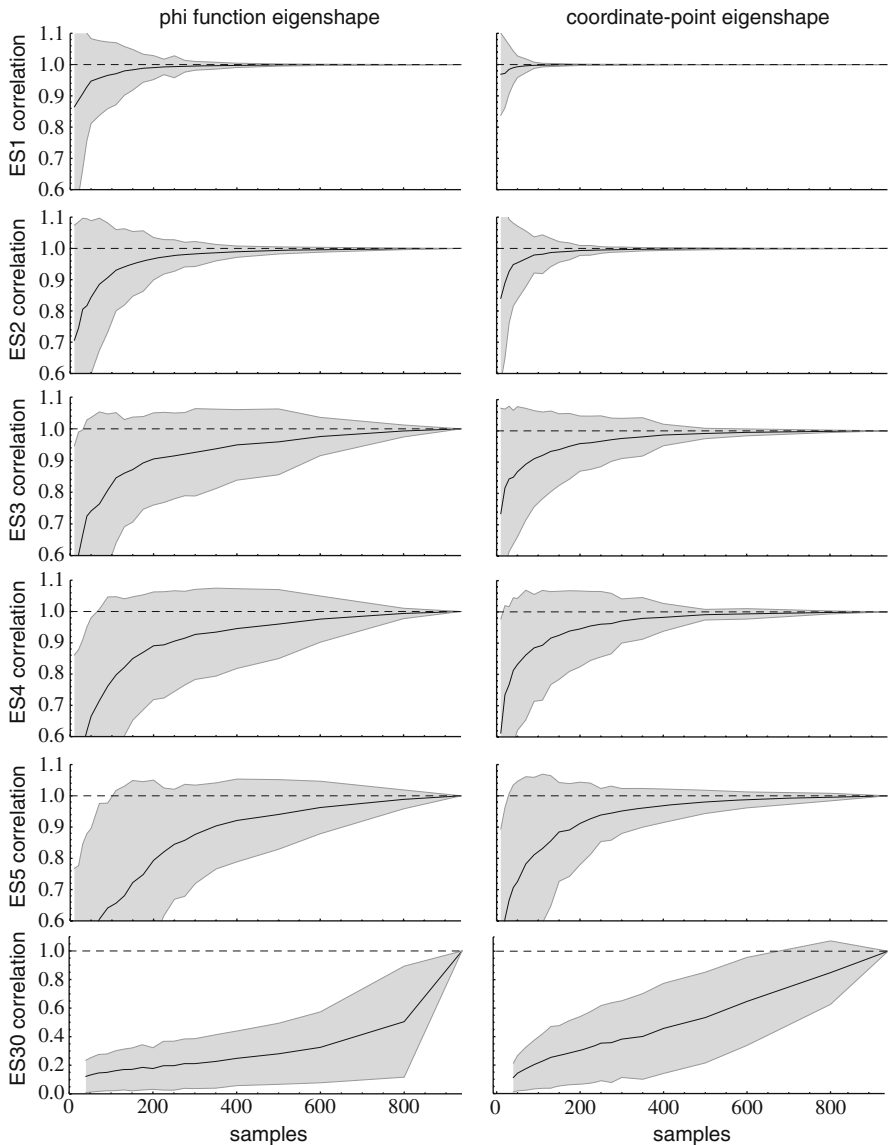
Fig. 3.8 Resampling for the analyses of raw outline data. Each plot shows Pearson's product-moment correlation coefficient for eigenscores between the subsampled analyses and the complete data set at a range of sample sizes. The black line is the mean correlation and the gray band is the 95% confidence interval for the mean. Note the change in vertical axis for ES30, since most of the mean correlations are below 0.6. ES30 starts with a sample size of 40, since there can be no more eigenshapes than the number of objects or points minus 1. The last two sample sizes are 800 and 937, so the ES30 phi-ES plot may paint an inaccurate picture for a high sample size
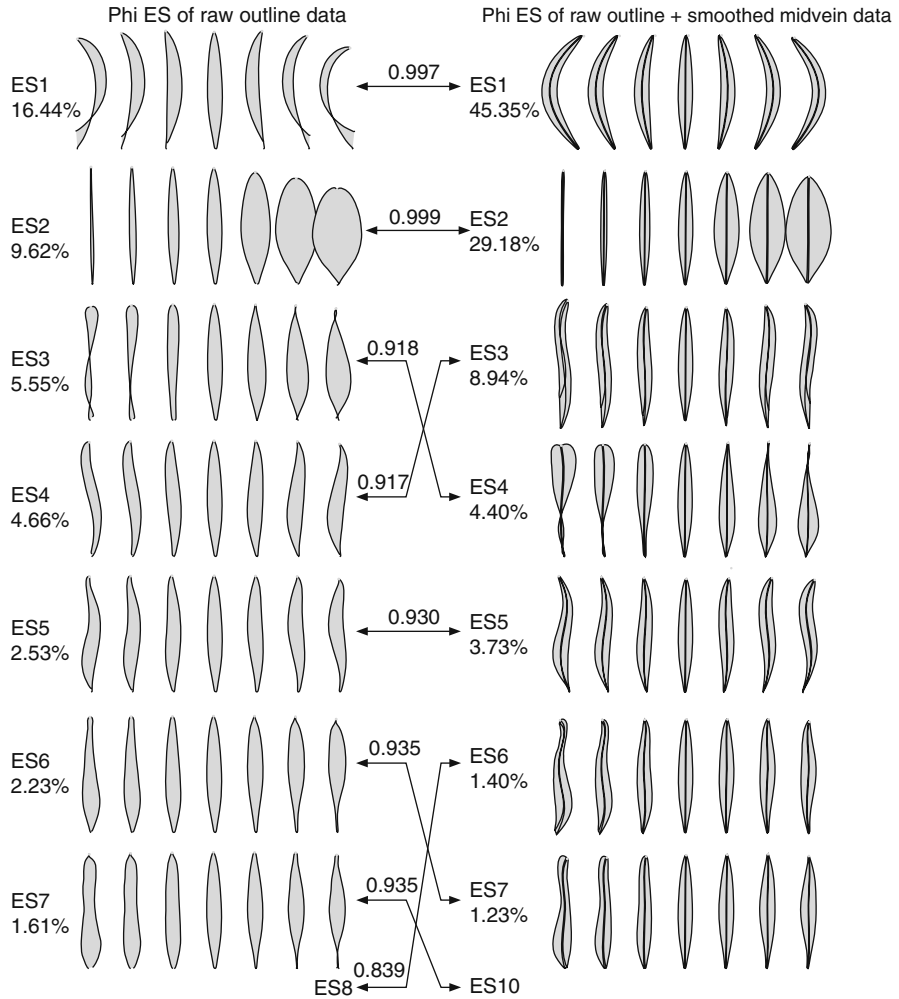
Fig. 3.9 Phi eigenshape analysis of raw leaf outlines (LA1) versus raw outlines with appended smoothed midveins (LA2). A landmark indicates the leaf apex. Percent variance explained is given for each axis. Pearson correlation coefficients for eigenscores between two analyses are shown where greater than 0.5. While the ordering and variance explained are slightly different, the first seven eigenshapes of each analysis are present in the other, with correlations higher than 0.8 in all cases. For all plots, the positive (high) end of the axis is to the right, though the direction does not matter

ordering of axes. The change in order between ES3 and ES4 was due to the added variance from midvein curvature; it was profound along ES3 (in LA2), and absent from ES4. Nevertheless, these two forms of outline variation were present in both analyses. ES5 appeared to be a similar form of curving variation, with a slightly lower correlation. This was presumably due to the slight additional variance seen
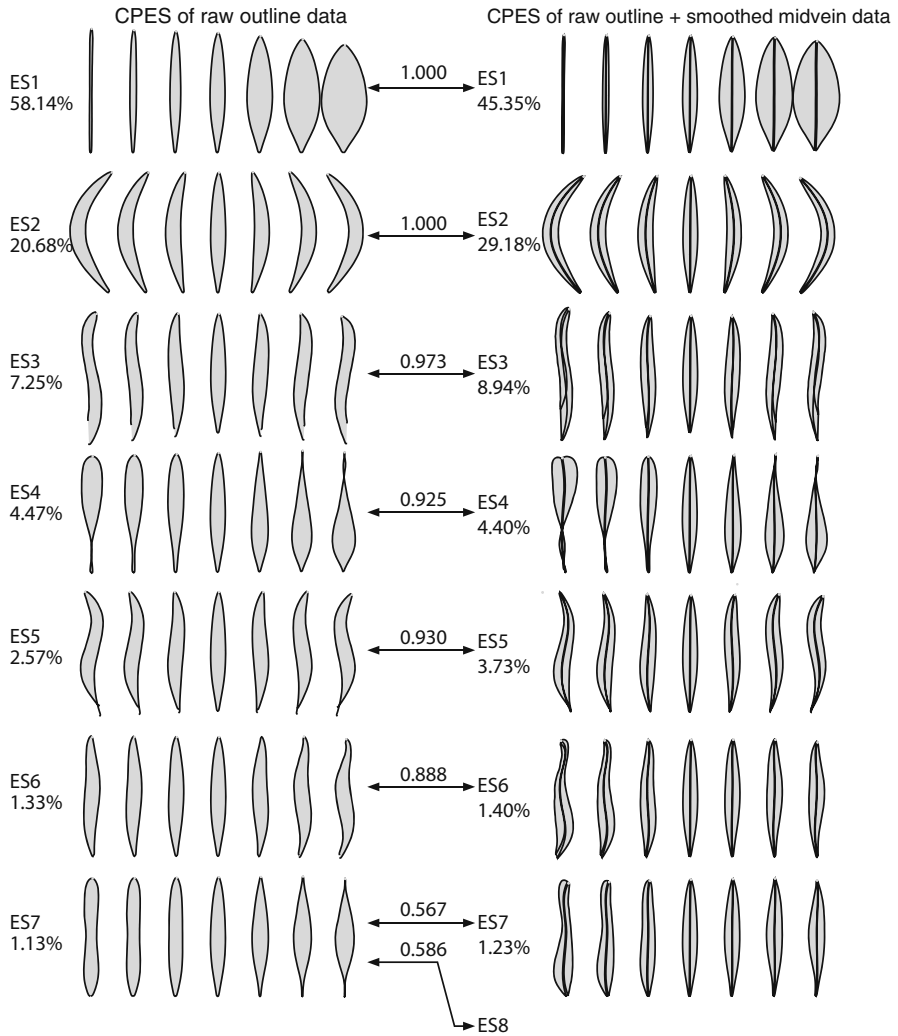
Fig. 3.10 Coordinate-point eigenshape analysis of raw leaf outlines (LA1) versus raw outlines with appended smoothed midveins (LA2). Ordering and percent variance explained are very similar for the first six eigenshapes. The seventh eigenshape in LA1 is correlated, albeit not very strongly, with ES8 of LA2, and the seventh eigenshape in LA2 appears to be a novel, asymmetric shape measure, as the original ES7 has split into two measures (ES8 in LA2 is similar to ES7, with slight curvature at both ends)

in the midvein along this axis, with a corresponding increase in variance explained, from 2.5% to 3.3%.

Individual axes in the phi analyses look very interpretable using qualitative terminology. Given the high correlations between the two analyses, it is sufficient to describe the models for the analysis of raw outlines. This will be useful for

comparison with the straightened leaf analysis. ES1 was clearly overall leaf curvature. The ends of the models were open, which is a common problem in analyses using phi functions, once which has been recognized since their original formulation in Zahn and Roskies (1972); this is one benefit to using coordinate-point eigenshape analysis, as will be seen below. ES2 appeared to be circularity (to use an existing term), or perhaps "ellipticity", since the variation is from a linear leaf to an elliptic leaf with slight asymmetry at the ends (rounded tip and an acute base). ES3 was a bit harder to interpret because of another phi-function artifact; here the leaf outline crosses over itself twice on the low end of the axis. Starting from the high end of the axis, it is easier to see that this axis describes a leaf wider at the tip versus base, or ovate to obovate variation. This has also been seen in eigenshape analysis of dicots (Krieger et al. 2007), but the pattern is much more subtle in fern leaves, which tend to be much longer relative to their width. ES4 and 5 appear to represent curvature, though ES4 also seems to have some asymmetry at the base. ES3 in LA2, highly correlated with ES4, was used to identify the pattern: it is a combination of curvature and linear-to-elliptic variation. ES6 was another form of ovate-to-obovate variation, more localized to the leaf ends. ES7 might be described as oblong to elliptic variation. However, these terms (see Fig. 3.1) don't seem to capture the very ends of the axis, with an oblong leaf that is pinched at the middle on one end, and an elliptic leaf that is pinched at the ends on the other.

The analysis of coordinate-point data had very similar results. The first six eigenshapes looked very similar between the two analyses, all with correlations greater than 0.88 (Fig. 3.10). Addition of the midvein has apparently allowed the eigenanalysis to split the original ES7 into two shape measures, each showing different patterns of curvature (the models for ES8 in LA2 are curved on both ends of the leaf).

Interpretation of the axes will again focus on the raw outline data (LA1). ES1 showed the same linear to elliptic pattern seen in the phi analyses, with a bit less asymmetry at the ends. This axis may be a bit closer to a true circularity axis, in that extending the models farther to the high end of the axis (past the highest scoring specimen) may generate circular or near-circular shapes. In the phi analyses, the strong tip-base asymmetry and slightly less rounded base would prevent that. ES2 was the same overall curvature axis, albeit reversed. ES3 looked less like a curvature axis and more like variation due to asymmetry in the extension of the lamina along the petiole (a pattern that is much clearer in the straightened leaf analysis). The remaining patterns were also seen in the phi analyses: ES4 was ovate-obovate variation, ES5 and 6 were curvature, and ES7 was oblong-elliptic variation.

The analysis including the midvein showed that ES5 included both curvature and slight ovate-obovate variation. ES7, which is appealing in its similarity to qualitative descriptors, was absent from LA2, having been split between two axes with additional asymmetry and curvature. ES7 in LA2 (shown in Fig. 3.10) had the same pattern as ES7 in LA1, with the addition of pronounced asymmetry at the base and slight curvature, and ES8 in LA2 had the addition of curvature at both ends, sharply curving to the right on the high end of the axis.

Comparison of Figs. 3.9 and 3.10 illustrates some of the differences between phi and coordinate-point eigenshape analysis. Contrast models for the first two eigenshapes: the CPES models were well-behaved, whereas the phi models had open boundaries (the leaf outline crosses over itself on both ends of ES1) and the midline extended beyond the edge of the leaf. The strange pattern seen in phi-ES1 relative to CP-ES2 is attributable to differences in information content between the two types of analysis. The issue here is not a simple case of information loss *per se*, as the interpolated outlines and midveins can be perfectly reconstructed from the phi functions. However, the decomposition of phi functions has less information to work with, as there are only four segment lengths (points are equally spaced within each of the four segments: the two sides of the outline and midvein) in the phi analysis, and the relative locations of outline and midvein points are specified by a single angle, where the two meet. All of this geometric information is present in the raw (aligned) coordinate data, which is used in the coordinate-point analysis.

Looking at the analyses of raw outline data, the curvature variation seen in phi-ES4 was very similar to CP-ES3. However, the geometric information present in the CP analysis, which specifies the relative location of the two ends of the lamina outline, makes it clear that this axis represents asymmetry at the base of the lamina, not simply waviness of the leaf. Though phi-ES3 and CP-ES4 both represented ovate-obovate variation, the CP models had fewer problems with the outline crossing itself. However, this is still present to a small extent on the extremes of the axis. This is because each of these axes is, ultimately, a mathematical entity contributing in a small way to the actual shape of the leaf. As they represent components of shape, and not the object shapes themselves, the models need not be biologically possible; as variation is modeled at or past the extent of specimens along an axis, the models can quickly become very strange. CP-ES9 (not shown) was very similar to phi-ES6, so this shape measure had not been lost. ES7 was also very similar between the two analyses, and this axis illustrates another difference: CP models tend to represent purer geometric variation. Note that phi-ES7 had asymmetry between the ends (particularly visible on the model at the left end of the axis), and slight left-right asymmetry, whereas CP-ES7 had very little. This is also apparent in a comparison of the coordinate-point axes ES1 and 4, and their phi counterparts.

Regardless of these minor differences, despite the two very different types of information (cumulative angular change versus Procrustes aligned coordinate data), the resulting shape measures are very similar. One major difference which is difficult to explain is the structure of the variance. In the phi analyses, there was a flat curve with a long tail: variance explained started at 16% for ES1 and slowly decayed. In the coordinate-point analyses, it started much higher, 58% in the analysis of raw data, and promptly dropped such that ES7 explained less of the variance than in the corresponding phi analysis. This may be because very small-scale variations contributed a large amount of variance in angle change, whereas for CPES the variation in location of coordinates was a tiny contribution to variation in the overall geometry. The contrast in variance structure between phi-ES and

CPES, with otherwise similar axes, supports the approach commonly used in eigen-shape analysis, to downplay the percent variance explained and instead focus on the eigenshapes themselves and their ability to describe biological features.

LA3 vs. LA1 and LA2 - Straightening

Results for the phi-ES analysis of straightened outlines are shown in Table 3.2 and Fig. 3.11. The first five eigenshapes in LA3, which correspond to symmetric varia-tion, were highly correlated with axes in both of the unstraightened analyses. ES6 and 7, exhibiting asymmetric variation, were more weakly correlated, and ES6 had no correlations greater than 0.5 with any of the first 25 axes in the analysis of raw outline data. In all cases, the straightened axes explained a greater proportion of the variance than their counterparts in LA1 and LA2, despite having variation due to curvature removed.

The amount of variance due to curvature is difficult to estimate, because the total variance changed very little between LA1 and LA3; it was much higher in LA2

Table 3.2 Relationships among the three phi eigenshape analyses. Axes present in the straightened analysis are shaded gray. The curvature axes in LA1—ES1, 4, and 5—are gone. Together, these represent 26.63% of the variance explained. ES8, 10, and 11 are also absent from LA3, but the models are less interpretable

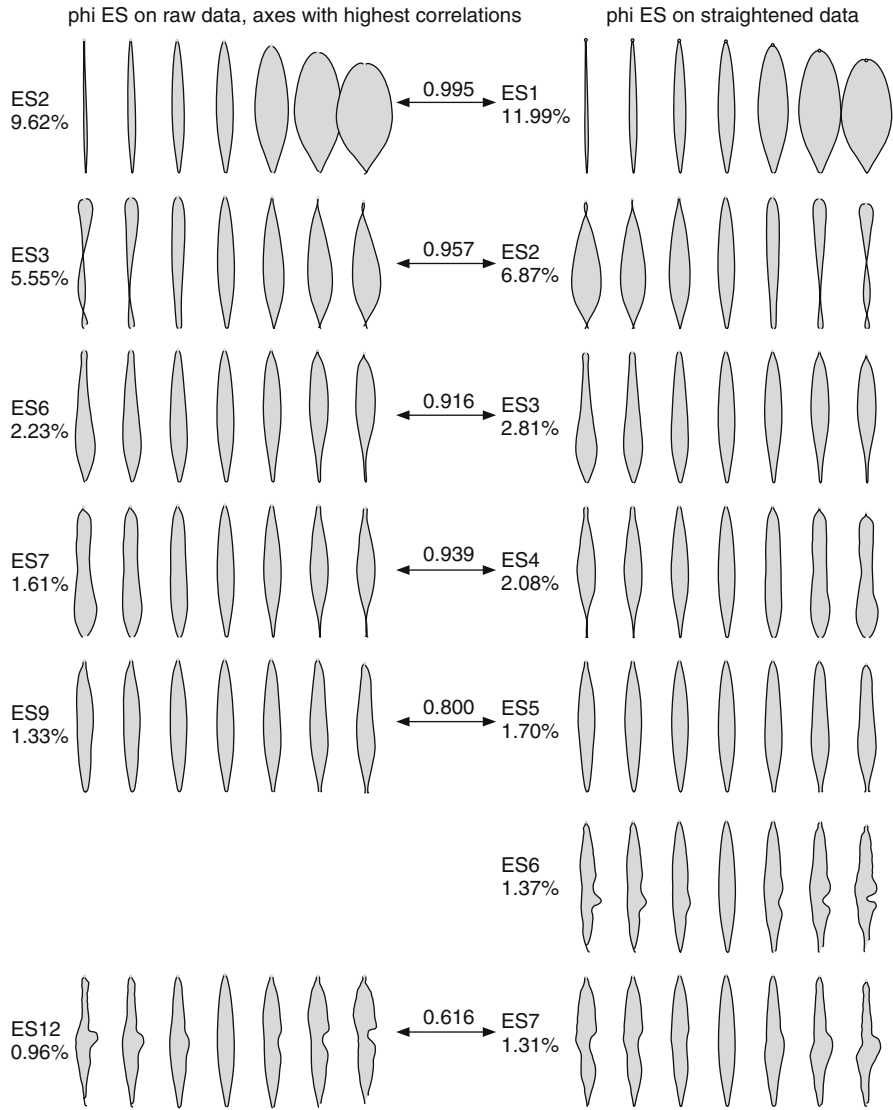| var. (%)	raw data	straightened	raw data + midvein	var. (%)
16.44	1	→0.995→ 1 ←0.993←	1	24.24
9.62	2	→0.957→ 2 ←0.974←	2	7.14
5.55	3	→0.916→ 3 ←0.946←	3	6.80
4.66	4	→0.939→ 4 ←0.867←	4	4.07
2.53	5	→0.800→ 5 ←0.904←	5	3.32
2.23	6	6 ←0.589←	6	1.87
1.61	7	→0.616→ 7 ←0.653←	7	1.68
1.53	8		8	1.52
1.33	9		9	1.39
1.17	10		10	1.14
0.99	11		11	0.98
0.96	12		12	0.95
			13	0.78
			14	0.71
			15	0.68

Fig. 3.11 Phi-ES on straightened leaves versus raw outlines. Each of the LA3 eigenshapes explains a slightly higher percent of the variance. Specimen score correlations agree with comparison of the models: the first five eigenshapes in LA3 are also seen in the analysis of raw outline data. For ES6 and 7, axes showing asymmetric variation, the relationships are much weaker

because of the addition of the midvein. However, it is possible to tally up the variance explained by curvature-related axes that are absent from LA3, as a rough estimate. In LA1, ES1, 4, and 5 were clearly curvature-related, though ES4 and 5 appeared to also contain asymmetry. Together they explained 23.63% of the variance. ES8, 10, and 11 were missing from LA3, but they explained very small

amounts of variance, and later axes tended to mix curvature and asymmetry. In LA2, ES1, 3, 5, and 6 were clearly curvature-related, though ES5 and 6 appeared to contain asymmetry. Together they explained 36.23% of the variance. These values were higher because of the contribution of the midvein; by definition, all of the variance contributed by the midvein is curvature.

The straightened leaf models looked very similar to the unstraightened leaf models, differing primarily in reduced asymmetry (e.g., ES2 and 3 in Fig. 3.11). Given that curvature, not asymmetry, was removed, the asymmetries in the LA1 models for ES3 and ES6 were likely due to sample-dependent covariance of curvature with shape. A larger sample size would reduce this effect. While ES5 was highly correlated with ES9 in the raw outline analysis, the removal of curvature made this axis much easier to interpret as localized ovate-obovate variation (Fig. 3.11).

Results for the CPES analysis of straightened outlines are shown in Table 3.3 and Fig. 3.12.

Table 3.3 Relationships among the three coordinate-point eigenshape analyses. Curvature axes from LA1—ES 2 and 5—are absent from LA3. These two axes represent 23.25% of the variance explained, nearly identical to the axes clearly representing curvature in phi LA1. ES3 in the analysis of raw data appears to represent curvature (Fig. 3.11), but it is highly correlated with ES2 of the straightened leaf analysis: asymmetry in the extension of petiole at the leaf base (Fig. 3.12). The pattern of higher correlations suggests that the analysis of raw leaves (LA1) outperformed the combined analysis (LA2) in identifying curvature-free shape measures. The straightened (LA3) ES6 is absent from LA2 (highest correlation 0.499), and ES4 has a much lower correlation with its LA2 counterpart

var. (%)	raw data	straightened	raw data + midvein	var. (%)
58.14	1	1.000 → 1 ← 0.999	1	45.35
20.68	2	0.836 2 ← 0.913	2	29.18
7.25	3	0.992 3 ← 0.919	3	8.94
4.47	4	0.972 → 4 ← 0.624	4	4.40
2.57	5	0.635 5 ← 0.596	5	3.73
1.33	6	0.909 6	6	1.40
1.13	7	0.564 → 7 ← 0.519	7	1.23
0.74	8		8	0.89
0.56	9		9	0.69
0.40	10		10	0.44
0.34	11		11	0.37
0.28	12		12	0.32
0.23	13		13	0.29
0.21	14		14	0.22
0.18	15		15	0.22

CPES on raw data, axes with highest correlations   CPES on straightened data
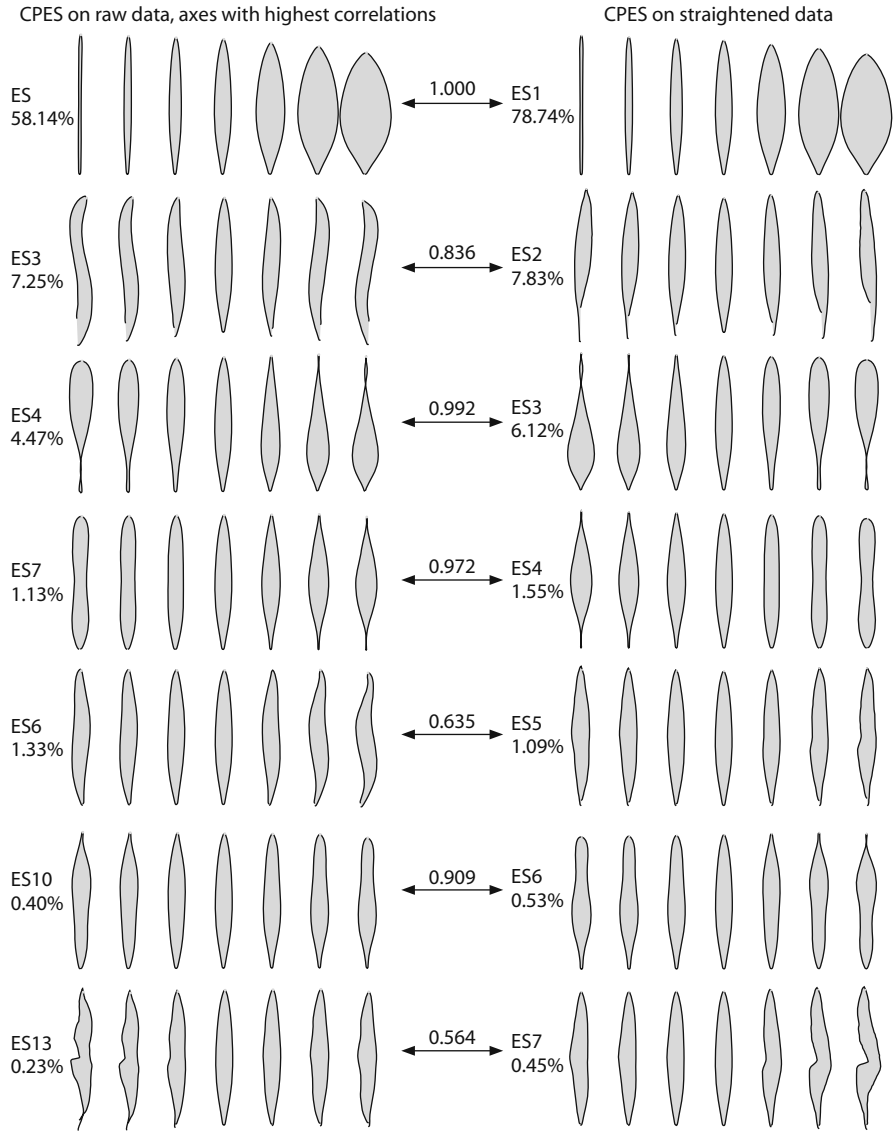

Fig. 3.12 CP-ES on straightened leaves versus raw outlines. Most of the LA3 eigenshapes explain a slightly higher percent of the variance. ES1, 3, 4, and 6 are nearly indistinguishable from their LA1 counterparts, with very high correlations. This is not true for ES2, an asymmetry axis, nevertheless highly correlated with ES3. For ES6 and 7, axes showing asymmetric variation, the relationships are much weaker. Asymmetry axes (ES5, 7) have much weaker correlations, with lower similarity between analyses

The results were less consistent among analyses than for phi-ES. The scores along the first three eigenshapes in LA3 were highly correlated with axes in both of the unstraightened analyses. ES1 and 3 corresponded to symmetric variation, and ES2 was asymmetry in the extension of the lamina. In both unstraightened analyses, ES3 was a mix of curvature and asymmetry. Nevertheless, it still explained a comparable amount of variance. ES4 in LA1 was strongly correlated with ES7 in the analysis of raw data, but weakly correlated with axes from LA2. This is presumably because, in LA2, this pattern of shape variation was split into two measures, with different types of asymmetry and curvature in each. ES6, localized ovate-obovate variation, was also strongly correlated with an axis in LA1 but not LA2. The relatively weak correlations between ES7 and ES13 (from both analyses) suggested that the original variables contain a considerable amount of curvature. In most cases, the straightened axes explained a greater proportion of the variance than their counterparts in LA1 and LA2, the exceptions being ES2, which explained both curvature and asymmetry before straightening, and ES5, another asymmetry measure that is weakly correlated with axes containing curvature.

The amount of variance due to curvature removed was comparable to the phi-ES analyses. In LA1, ES2 and 5 were clearly curvature-related, though ES5 appeared to also contain localized ovate-obovate variation. Together they explained 23.25% of the variance. ES8, 9, 11, and 12 were also missing from LA3, but together they explained less than 2% of the variance. In LA2, ES2 and 5 were clearly curvature-related, though again with ovate-obovate variation on ES5. Together they explained 32.81% of the variance.

The models looked virtually identical to the unstraightened models; compare models for ES1, 3, 4, and 6 with those in LA1, (Fig. 3.12). ES2 appeared to correspond to a single asymmetry character: basal lamina extent. Though correlated with ES6 in both unstraightened analyses, ES5 appeared to be a novel asymmetry character, as well. Though it was correlated with ES13 in both analyses, the relatively weak correlations suggested that ES7 was a somewhat novel asymmetry character, now that curvature has been removed.

Discussion

The goal of this study was to determine the best way to extract shape measures from leaf outline data when the dataset contains a significant percentage of variation due to curvature. Eigenshape analysis appears to be well-suited to extracting symmetric shape measures from such datasets, and the results of this study provide guidance for the correct approach in undertaking such an analysis.

Will a sufficiently large sample size allow the eigenanalysis to identify and isolate curvature? The answer seems to be yes. The necessary sample size was roughly 200 for the coordinate-point eigenshape analysis, and 300 specimens for phi-ES. These values depend on the specific dataset being analyzed, both on similarity of the objects and on patterns of shape variation in the sample. However, randomization

can be used to determine whether individual axes have converged on a stable morphospace, or whether additional specimens are likely to change the eigenshape, hence the nature of the shape measurement. Given the high amount of curvature in this dataset, it is reasonable to expect even better results for a typical leaf dataset. High levels of asymmetry can be expected to have similar effects to high levels of curvature.

One question that arises in morphometric analyses is how even the sampling should be. For the majority of herbaria visited for this study, all sheets were examined for each species, and all sheets with material that appeared usable were photographed. Leaf damage and number of collections for rare species severely limited data collection in some cases, resulting in sample sizes from 11 to 100. The potential effects of skewed sample size are a change in the sample mean shape and changes in the orientation of individual eigenvectors. However, the benefit of broader sampling (more leaves in some species, where possible) appeared to outweigh the cost of uneven sampling: the uneven sampling did not appear to negatively impact construction of the empirical morphospace. This might manifest as outliers that are distorting the orientation of principal component axes. The only place this was seen was along ES2 in the coordinate point analysis of straightened leaves (Fig. 3.12), because a small number of leaves had extreme asymmetry in the extension of lamina along the petiole. This seemed like a real pattern, where a long-tailed distribution would be expected, so the outliers were not removed from the analysis. The outliers may have inflated the correlation between this axis and the corresponding axes in the other analyses, which seemed to have a mix of curvature and asymmetry. In general, such outliers are easy to identify in plots of specimen scores, at which point their affects on the orientation of the eigenvectors can be assessed. This procedure should be performed with an eye on the biology: are the resulting eigenvectors (eigenshapes) biologically meaningful, and are the outliers simply the extremes of a continuum or are their shapes truly unrepresentative of the rest of the sample?

Should a curvature reference be appended, to allow the eigenanalysis to identify and isolate curvature? Surprisingly, the answer is, mostly, no. The models including a midvein made it easier to identify the mixing of asymmetry and curvature, but the two sources of variance were still mixed. In the CPES analyses, the raw outlines actually generated nicer shape measures, closer to those in the straightened analysis. Specifically, localized ovate-obovate variation is present in both, but has been lost as a character when the midvein is added. Overall, nearly all of the characters present in the straightened leaf analyses were present in the analyses of raw outline data.

It seems counterintuitive that the eigenanalysis could not segregate asymmetry and curvature when it was given the midvein. However, it may be that even a sample size of 937 was inadequate for a dataset with roughly 23% of the variance due to curvature. This is a fairly large sample size, so it may simply be unrealistic to develop datasets where this can be used. Collecting and processing the additional curve data is time consuming, and it introduces additional variance. Therefore, it seems like appending a reference curve is not worth the effort, unless it is to be used for straightening.

Should outline data be straightened, when a reference curve is available? The two previous answers suggest that the answer is no, but this depends on the questions being asked. If the focus is to extract symmetric shape measures (such as the majority illustrated in Fig. 3.1), an analysis of raw data should be adequate; curvature axes can be identified easily and excluded from further statistical analysis. For the fern dataset, there were four types of symmetric shape variation identified in the straightened analysis: linear-elliptic, ovate-obovate, oblong-elliptic, ovate-obovate localized to smaller regions of the leaf. These were identified equally well in the raw and straightened analyses, with correlations of scores between the two greater that 0.9 in all cases. The scores are the measurements made on these shape variables, so the high correlation shows that the same measurements are being made, with or without straightening. When such a set of measurements can answer the biological questions being asked, refraining from straightening can save a considerable amount of effort, in digitizing—unlike the outline, for the midvein automating data collection was not possible, because the midveins were both hard to see and often appeared discontinuous, when covered with scales—and analysis, checking the outlines for algorithmic artifacts after straightening is time intensive.

When asymmetric characters are of interest, straightening is required. An outline analysis cannot distinguish asymmetry from curvature without a reference curve, since the two are necessarily confounded in the outline. The results above suggest that simply including the reference in the analysis is not adequate; instead it should be used to adjust the curved outline data. Asymmetric characters in this study included a notch on one side of the leaf (e.g., Fig. 3.6b), or a bulge midway along one side of the leaf. These features are both present in *Pleopeltis*, and extracting them in this way would allow them to be tested for taxonomic information content. The notch character was present only in the straightened leaf analyses. The bulge character was present in the analyses of raw data, albeit much farther down the list (ES12/13 versus ES7), but the correlation coefficients between the two were relatively low, and in the raw analyses there appeared to be a mix of asymmetry and curvature.

Shape Measures

As described above, there were four symmetric shape variables consistently identified: linear-elliptic, ovate-obovate, oblong-elliptic, and localized ovate-obovate variation. The "oblong-elliptic" elliptic axis may represent a novel pattern of variation, since the extremes of the axis were instead dumbbell shaped on the oblong end, and elliptic with highly acute tip and base on the other. It is highly likely that there are terms already in use, specific to both of these shapes; however, they have probably not been recognized as the ends of a vector through a leaf sample mean shape.

A notable finding is that there were no axes representing classic tip or base shape variation, encompassed by the terms in Fig. 3.1. There was nothing inherent in the technique to prevent this being the case: eigenshape analysis on a broader sample of flowering plant leaves showed that tip shape varied independently of overall leaf

shape (Krieger et al. 2007), perhaps because of the prevalence of extended drip tips in tropical dicots. ES2 for the straightened, CPES analysis (Fig. 3.12) shows variation only at base of the leaf, in the form of asymmetric extension of the lamina down the petiole. All other eigenshapes show covariance between base, tip, and overall shape. This may reflect this particular sample of ferns; a combination of broad sampling (to capture cases where covariation differs among species) and within-species sampling (to capture dissociation of characters within species) is required to determine whether tip, base, and overall shape covary across the pteridophytes. However, it does suggest that scoring a leaf for each of these features separately may lead to analysis of strongly correlated measurements.

The eigenshapes generated in this study have the benefit of being statistically defensible shape measures: they have been constructed to be uncorrelated. Nevertheless, they may not be attractive to systematists. In that case, the empirical approach used here could be used to develop theoretical shape measures, at the potential cost of orthogonality among the measurements. For example, the linear-elliptic eigenshape could be adjusted to be a geometrically pure linear-circular shape, and specimens can be projected onto that axis to generate a set of scores (measurements). This can be done for any of the shapes, to make them more similar to existing qualitative terms. Starting with eigenshapes increases the chance that the measures will be minimally correlated, and it also provides an opportunity to discover novel measures of shape, like the "dumbbell/oblong" to "pointy elliptic" vector, or the three axes corresponding to different types of ovate-obovate variation. This type of approach seems superior to building a purely theoretical morphospace (where axes are determined *a priori* through some application of theory), in as much as this *empirical* morphospace was created from measurements of actual leaves and the covariance structure of a real dataset.

Measuring Curvature

Mathematical straightening of leaves worked very well, eliminating roughly 23% of the variance in this dataset, and improved interpretability of key eigenshape axes. This doesn't address the question of whether curvature should be treated as an undesirable source of variation, to be removed from analyses where possible. The few past studies of leaf shape that have mentioned curvature have treated it as an undesirable feature, usually eliminating overly curved leaves. Efforts have not been made to look for a biological signal in curvature, which might exist in cases, e.g., where plants adjust phyllotactic leaf orientation. A metric of curvature could be used to test this type of observation. The predominant linear approaches to leaf measurement would at best weakly record curvature and could not be trusted to identify an interpretable pattern. However, it should be relatively straightforward to expand upon the leaf straightening methods presented here to extract a measure of total curvature. This could be in one of two forms. The first is related to classic linear measurements of curvature: the total deflection of the midvein from a straight line. The original and straightened leaves would need to be aligned before computation

of the deflection distances; there may be a better choice than Procrustes superimposition, depending on the biological question. The second approach would be to perform the eigenanalysis on the curved data, and use scores along curvature axes as the metrics of curvature. In this way, measurements may be recorded for different types of curvature (e.g., ES1, 4, and 5 in the phi-ES analysis of raw outlines, Fig. 3.9), and each of the measurements is scaled according to sample variance structure.

Phi-ES and CPES

The goal of this study was not to compare phi-based and coordinate point eigenshape analysis, but several differences were apparent. CPES models overcome the longstanding problem with phi functions: the failure of the ends of closed curves to join up. CP models were generally more geometrically regular; for example, in the phi analysis the linear-elliptic models were asymmetric in terms of tip versus base shape. Of course, both are correct pictures of the covariance structure of the dataset; the phi analysis favored angular change over coordinate location, and this asymmetry is the result. CPES was able to identify asymmetry in extension of the lamina at the base of the leaf, which seemed to be completely absent from all phi analyses. This was a surprising result, given that this was nearly 8% of the variance explained, the second largest eigenvalue in the CPES analysis of straightened leaves.

   The straightening of simple leaves is not yet fully automated for phi-based eigenshape analysis. The outline on the outer edge of a curve has more points than the inner edge, so straightening compresses many points into a small area, occasionally resulting in loops that have major impacts on the phi functions. There is no easy approach to the automated correction of these features. Repeated smoothing of the straightened outlines using a 10-point average was only effective for 2 of the 14 leaves with loops, and it was ultimately not used, out of concern that it would change leaf shape in other ways. A brute-force test of whether individual segments cross one another should work, though it would be computationally expensive. Despite the negative effects on the amplitudes of the phi functions, the loops are an indication of an otherwise healthy algorithm: it conserves the order of points along the outline. Other straightening methods may sample outline points out of order (e.g., the midvein-normal approach in West and Noble 1984) or reassemble the outline out of order. Using CPES, the loops can safely be ignored.

Conclusion

This study supports adjusting curved specimens where possible to obtain better measures of asymmetry, while also showing that this procedure is probably not necessary where asymmetry is not of interest. Eigenanalysis is able to identify most of the symmetric variation even in datasets with a large percentage of highly curved

objects. Most biological objects will not provide such a convenient guide to straightening as the midvein; this is true even for leaves, and these results suggest that eigenanalysis can be very valuable for such systems.

Both phi-based and coordinate point eigenshape are able to extract shape measures consistent with qualitative shape terminology, though it appears that base and tip shape do not vary independently from overall shape. An option worth further exploration is to create the empirical space using both original and reflected specimens. This should decrease the impact of small asymmetries on the construction of eigenshape axes. Another option is to include symmetrized versions of leaves in the decomposition (split each leaf in two, reflect each half to form two perfectly symmetric leaves). This may extract additional types of symmetric variation.

Asymmetry and curvature are important components of leaf form, both problematic in their ability to obscure other characters and potentially informative for measuring genetic and environmental effects on leaf development. Eigenshape analysis and straightening provide tools to precisely quantify these sources of variance.

References

Allen EG (2006) New approaches to Fourier analysis of ammonoid sutures and other complex, open curves. Paleobiology 32: 299–315

Avery GS Jr (1933) Structure and development of the tobacco leaf. American Journal of Botany 20(9): 565–592

Belongie S, Malik J (2000) Shape matching and object recognition using shape contexts. IEEE transactions on Pattern Analysis and Machine Intelligence 24(24): 509–522

Belongie S, Malik J, Puzicha J (2000) Shape context: a new descriptor for shape matching and object recognition. In Leen TK, Dietterich TG, Tresp V (Eds.) Advances in Neural Information Processing Systems, volume 13, 831–837 pp

Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(24): 509–522

Bensmihen S, Hanna AI, Langlade NB, Micol JL, Bangham A, Coean ES (2008) Mutational spaces for leaf shape and size. Human Frontier Science Program Journal 2(2): 110–120

Black-Samuelsson S, Whiteley RE, Junzhan G (2003) Growth and leaf morphology response to drought stress in the riparian broadleaved tree, Ulmus laevis (Pall.). Silvae Genetica 52(5–6): 292–299

Bookstein FL (1986) Size and shape spaces for landmark data in two dimensions. Statistical Science 1: 181–222

Bookstein FL (1997) Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis 1(3): 225–243

Bookstein FL (2002) Creases as morphometric characters. In MacLeod N, Forey PL (Eds.) Morphology, shape and phylogeny. Systematics Association Special volume Series 64, Taylor and Francis, London, 139–174 pp

Bookstein FL, Strauss RE, Humphries JM, Chernoff B, Elder RL, Smith GR (1982) A comment upon the uses of Fourier methods in systematics. Systematic Zoology 31(1): 85–92

Boyce CK, Knoll AH (2002) Evolution of developmental potential and the multiple independent origins of leaves in Paleozoic vascular plants. Paleobiology 28(1): 70–100

Bylesjo M, Segura V, Soolanayakanahally RY, Rae AM, Trygg J, Gustafsson P, Jansson S, Street NR (2008) LAMINA: a tool for rapid quantification of leaf size and shape parameters. BMC Plant Biology 8: 82

Campbell RD (1996) Describing the shapes of fern leaves: a fractal geometric approach. Acta Biotheoretica 44: 119–142

Cannon CH, Manos PS (2001) Combining and comparing morphometric shape descriptors with a molecular phylogeny: the case of fruit type evolution in Bornean *Lithocarpus* (Fagaceae). Systematic Biology 50(6): 860–880

Caumul R, Polly PD (2005) Phylogenetic and environmental components of morphological variation: skull, mandible, and molar shape in marmots (*Marmota*, Rodentia). Evolution 59(11): 2460–2472

Chi Y-T, Chien C-F, Lin T-T (2003) Leaf shape modeling and analysis using geometric descriptors from Bezier curves. Transactions of the ASAE 46(1): 175–185

Dale MB, Groves RH, Hull VJ, O'Callaghan JF (1971) A new method for describing leaf shape. New Phytologist 70: 437–442

Dancik BP, Barnes BV (1974) Leaf diversity in yellow birch. Canadian Journal of Botany 52: 2407–2414

Dilcher DL (1974) Approaches to the identification of angiosperm leaves. The Botanical Review 40(1).

Ehrlich R, Pharr RB Jr, Healy-Williams N (1983) Comments on the validity of Fourier descriptors in systematics: a reply to Bookstein et al. Systematic Zoology 32(2): 202–206

Elewa AMT (2003) Morphometric studies on three ostracod species of the genus Digmocythere Mandelstam from the Middle Eocene of Egypt. Palaeontologia Electronica 6(2): 11

Farley RA, McNeilly T (2000) Diversity and divergence in *Cistus salvifolius* (L.) populations from contrasting habitats. Hereditas 132: 183–192

Gage E, Wilkin P (2008) A morphometric study of species delimitation in Sternbergia lutea (Alliaceae, Amaryllidoideae) and its allies S. sicula and S. greuteriana. Botanical Journal of the Linnean Society 158: 460–469

Guralnick RP (2005) Combined molecular and morphological approaches to documenting regional biodiversity and ecological patterns in problematic taxa: a case study in the bivalve group *Cyclocalyx* (Sphaeriidae, Bivalvia) from western North America. Zoologica Scripta 34(5): 469–482

Hageman W (1992) The relationship of anatomy to morphology in plants: a new theoretical perspective. International Journal of Plant Sciences 153(3): S38–S48

Hammel M, Prusinkiewicz P, Wyvill B (1992) Modelling compound leaves using implicit contours. In Kunii TL (Ed.) Visual computing: integrating computer graphics with vision. Springer-Verlag, Tokyo, 119–212 pp

Heggie M, Zodrow EL (1994) Fractal Lobatopterid frond (Upper Carboniferous Marattialean tree fern). Palaeontographica. Abteilung B: Paläophytologie 232: 35–57

Hickey LJ (1973) Classification of the architecture of dicotyledonous leaves. American Journal of Botany 60(1): 17–33

Hickey LJ (1979) A revised classification of the architecture of dicotyledonous leaves. In Metcalf CR, Chalk L (Eds.) Anatomy of the dicotyledons, Second Edition. Clarendon Press, Oxford

Hooper EA (1994) Biosystematic analysis of the *Pleopeltis macrocarpa* complex in the neotropics. Ph.D. thesis, University of Kansas

Huff PM, Wilf P, Azumah EJ (2003) Digital future for paleoclimate estimation from fossil leaves? Preliminary results: PALAIOS 18: 266–274

Iwatsuki K (1990) Hymenophyllaceae. In Kramer KU, Green PS (Eds.) volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin

Jensen RJ (1990) Detecting shape variation in oak leaf morphology: a comparison of rotational-fit methods. American Journal of Botany 90: 333–338

Jensen RJ, Ciofani KM, Miramontes LC (2002). Lines, outlines, and landmarks: morphometric analysis of leaves of *Acer rubrum*, *Acer saccharinum* (Aceraceae) and their hybrid. Taxon 51: 475–492

Jensen RJ, Hokanson SC, Isebrands JG, Hancock JF (1993) Morphometric variation in oaks of the Apostle Islands in Wisconsin: evidence of hybridization between *Quercus rubra* and *Q. ellipsoidalis* (Fagaceae). American Journal of Botany 80(11): 1358–1366

Jones CS (1992) Comparative ontogeny of a wild Cucurbit and its derived cultivar. Evolution 46(6): 1827–1847

Jones CS (1993) Heterochrony and heteroblastic leaf development in two subspecies of *Cucurbita argyrosperma* (Cucurbitaceae). American Journal of Botany 80: 778–795

Kaplan DR (1992) The relationship of cells to organisms in plants: problem and implications of an organismal perspective. International Journal of Plant Sciences 153(3): S28–S37

Kaplan DR (2001) Fundamental concepts of leaf morphology and morphogenesis: a contribution to the interpretation of molecular genetic mutants. International Journal of Plant Sciences 162(3): 465–474

Kelly DC, Bralower TJ, Zachos JC (2001) On the demise of the early paleogene *Morozovella velascoensis* lineage: terminal progenesis in the planktonic foraminifera. PALAIOS 16 (5): 507–523

Kincaid DT, Schneider RB (1983) Quantification of leaf shape with a microcomputer and Fourier transform. Canadian Journal of Botany 61: 2333–2342

Kott LS, Britton DM (1982) A comparative study of sporophyte morphology of the three cytotypes of Polypodium virginianum in Ontario. Canadian Journal of Botany 60: 1360–1370

Kramer KU (1990a) Cyatheaceae. In Kramer KU, Green PS (Eds.) The families and genera of vascular plants volume I: Pteridophytes and gymnosperms. Springer-Verglag, Berlin.

Kramer KU (1990b) Gleicheniaceae. In Kramer KU, Green PS (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K, (Ed.) Springer-Verglag, Berlin.

Kramer KU (1990c) Marsileaceae. In Kramer KU, Green PS, (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin.

Kramer KU (1990d) Osmundaceae. In Kramer KU, Green PS, (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin.

Kramer KU (1990e) Schizaeaceae. In Kramer KU, Green PS (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin.

Krieger JD (2007) Quantification and analysis of variation in leaf shape: climate signals in dicot leaves and evolution of leaf form in *Pleopeltis* (Polypodiaceae). Doctoral dissertation. University of Colorado, Boulder

Krieger JD, Guralnick RP, Smith DS (2007) Generating empirically determined, continuous measures of leaf shape for paleoclimate reconstruction. PALAOIS 22: 212–219

Kuhl FP, Giardina CR (1982) Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18: 236–258

Langlade NB, Feng X, Dransfield T, Copsey L, Hanna AI, Thebaud C, Bangham A, Hudson A, Coen E (2005) Evolution through genetically controlled allometry space. Proceedings of the National Academy of Sciences (USA) 102(29): 10221–10226

Leaf Architecture Working Group (1999) Manual of leaf architecture: morphological description and categorization of dicotyledonous and net-veined monocotyledonous angiosperms, Smithsonian Institution Press, Washington, DC, http://www.yale.edu/peabody/ collections/pb/pb_mla.html

Lele S (1993) Euclidean distance matrix analysis (EDMA): estimation of mean form and mean form difference. Mathematical Geology 25: 573–602

Lele S, Richtsmeier JT (1991) Euclidean distance matrix analysis: a coordinate-free approach for comparing biological shapes using landmark data. American Journal of Physical Anthropology 87: 415–427

Lellinger DB (2002) A modern multilingual glossary for taxonomic pteridology. In Pteridologia, volume 3, American Fern Society, USA, 263 p

Liow LH (2006) Do deviants live longer? Morphology and longevity in trachyleberidid ostracodes. Paleobiology 32(1): 55–69

Lohmann GP (1983) Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. Mathematical Geology 15: 659–672

Lohmann GP, Schweitzer PN (1990) On eigenshape analysis. In Rohlf FJ, Bookstein FL (Eds.) Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology, Ann Arbor, 145–166 p

MacLeod N (1995) Testing morphometric data for phylogenetic and functional covariance. Journal of Vertebrate Paleontology 11, Supplement No. 3: 41A–42A

MacLeod N (1999) Generalizing and extending the eigenshape method of shape space visualization and analysis. Paleobiology 5: 107–138

MacLeod N (2001) Landmarks, localization, and the use of morphometrics in phylogenetic analysis. In Adrian JM, Edgecombe GD, Lieberman BS (Eds.) Fossils, phylogeny, and form, Kluwer Academic/Plenum Publishers, New York, 197–233 p

MacLeod N (2002a) Geometric morphometrics and geological shape-classification systems. Earth-Science Reviews 59: 27–47

MacLeod N (2002b) Phylogenetic signals in morphometric data. In MacLeod N, Forey PL (Eds.) Morphology, shape, and phylogeny. 100–138 p

MacLeod N (2008) Understanding morphology in systematic contexts: 3D specimen ordination and 3D specimen recognition. In Wheeler Q (Ed.) The new taxonomy, CRC Press, Taylor & Francis Group, London, 143–210 pp

MacLeod N, Rose KD (1993) Inferring locomotor behavior in Paleogene mammals via eigenshape analysis. American Journal of Science 293-A: 300–355

McClain CR (2005) Bathymetric patterns of morphological disparity in deep-sea gastropods from the Western North Atlantic Basin. Evolution 59(7): 1492–1499

McClain CR, Johnson NA, Rex MA (2004) Morphological disparity as a biodiversity metric in lower bathyal and abyssal gastropod assemblages. Evolution 58: 338–348

McLellan T (1990) Development of differences in leaf shape in *Begonia dregei* (Begoniaceae). American Journal of Botany 77(3): 323–337

McLellan T (2000) Geographic variation and plasticity of leaf shape and size in *Begonia dregei* and *B. homonyma* (Begoniaceae). Botanical Journal of the Linnean Society 132: 79–95

McLellan T, Dengler NG (1995) Pattern and form in repeated elements in the development of simple leaves of *Begonia dregei*. International Journal of Plant Sciences 156: 581–589

McLellan T, Endler JA (1998) The relative success of some methods for measuring and describing the shape of complex objects. Systematic Biology 47: 264–281

Meade C, Parnell J (2003) Multivariate analysis of leaf shape patterns in Asian species of the Uvaria group (Annonaceae). Botanical Journal of the Linnaean Society 143: 231–242

Menesatti P, Costa C, Paglia G, Pallotino F, D'Andrea S, Rimatori V, Aguzzi J (2008) Shape-based methodology for multivariate discrimination among Italian hazelnut cultivars. Biosystems Engineering 101(4): 417–424

Mickel JT, Beitel JM (1988) Pteridophyte flora of Oaxaca, Mexico. The New York botanical garden, New York, 568 p

Monteiro LR, Bordin B, Reis SFd (2000) Shape distances, shape spaces and the comparison of morphometric methods. Trends in Ecology and Evolution 15(6): 217–220

Mundermann L, MacMurchy P, Pivovarov J, Prusinkiewicz, P (2003) Modeling lobed leaves. Proceedings of the Computer Graphics International, 2003: 60–65

Murakami NS, Nogami S, Watanabe M, Iwatsuki K (1999) Phylogeny of Aspleniaceae inferred from rbcL nucleotide sequences. American Fern Journal 89(4): 232–243

Neto JC, Meyer GE, Jones DD, Samal AK (2006) Plant species identification using Elliptic Fourier leaf shape analysis. Computers and Electronics in Agriculture 50: 121–134

Niklas KJ (1978) Morphometric relationships and rates of evolution among Paleozoic vascular plants. In Hecht MK, Steere WC (Eds.) Evolutionary biology, volume 11, Plenum Publishing Co., New York

Niklas KJ (1994a) Morphological evolution through complex domains of fitness. Proceedings of the National Academy of Sciences (USA) 91: 6772–6779

Niklas KJ (1994b) Simulation of organic shape: the roles of phenomenology and mechanism. Journal of Morphology 219: 243–246

Niklas KJ (1999) Evolutionary walks through a land plant morphospace. Journal of Experimental Botany 50(330): 39–52

Norris RD, Corfield RM, Cartlidge J (1996) What is gradualism? Cryptic speciation in globorotaliid foraminifera. Paleobiology 22(3): 386–405

Ollson A, Nybom H, Prentice HC (2000) Relationships between nordic dogroses (Rosa L. sect. Caninae, Rosaceae) assessed by RAPDs and elliptic Fourier analysis of leaflet shape. Systematic Botany 25(3): 511–521

Paler MH, Barrington DS (1995) The hybrid Cystopteris fragilis x C. tenuis (Dryopteridaceae) and the relationship between its tetraploid progenitors. Systematic Botany 20(4): 528–545

Plotze RD, Falvo M, Padua JG, Bernacci LC, Vieira MLC, Oliveira GCX, Bruno OM (2005) Leaf shape analysis using the multiscale Minkowski fractal, a new morphometric method: a study with *Passiflora* (Passifloraceae). Canadian Journal of Botany 83: 287–301

Polly PD (2003) Paleophylogeography: The tempo of geographic differentiation in marmots (*Marmota*). Journal of Mammalogy 84(2): 369–384

Polly PD, MacLeod N (2006) Locomotion in fossil Carnivora: an application of the eigensurface method for morphometric analysis of 3D surfaces. Palaeontologica Electronica 11(2): 13

Premoli AC (1996) Leaf architecture of South American *Nothofagus* (Nothofagaceae) using traditional and new methods in morphometrics. Botanical Journal of the Linnean Society, 121: 25–40

Prusinkiewicz P, Hammel M, Hanan J, Mech R (1996) L-systems: from the theory to visual models of plants. In Michalewicz (Ed.) Proceedings of the 2nd CSIRO Symposium on Computational Challenges in Life Sciences, 32 p

Prusinkiewicz P, Lindenmayer A (1990) The algorithmic beauty of plants. Springer-Verlag, New York

Pryer KM, Hearn DJ (2008) Evolution of leaf form in marsileaceous ferns: evidence for heterochrony. Evolution 63(2): 498–513

Ranker TA, Smith AR, Parris BS, Geiger JMO, Haufler CH, Straub SCK, Schneider H (2004) Phylogeny and evolution of grammitid ferns (Grammitidaceae): a case of rampant morphological homoplasy. Taxon 53(2): 414–428

Ray TS (1990) Application of "eigenshape" analysis to second order leaf shape ontogeny in *Syngonium podophyllum* (Araceae). In Rohlf FJ, Bookstein FL, (Eds.) Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology, Ann Arbor, 201–213 pp

Ray TS (1992) Landmark eigenshape analysis: homologous contours: leaf shape in *Syngonium* (Araceae). American Journal of Botany 79: 69–76

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. In Marcus LF, Bello E, Garcia-Valdecasas A (Eds.) Contributions to morphometrics, Madrid, Museo Nacional de Ciencias Naturales, 131–159 pp

Rohlf FJ (1996) Morphometric spaces, shape components, and the effects of linear transformations. In Marcus LF, Corti M, Loy A, Naylor G, Slice DE (Eds.) Advances in Morphometrics: Proceedings of the 1993 NATO Advanced Studies Institute on Morphometrics in Il Ciocco, Italy. Plenum, New York, 117–129 pp

Rohlf FJ (2000a) On the use of shape spaces to compare morphometric methods. Hystrix 11: 9–25

Rohlf FJ (2000b) Statistical power comparisons among alternative morphometric methods. American Journal of Physical Anthropology 111: 463–478

Rohlf FJ, Archie JW (1984) A comparison of Fourier methods for the description of wing shape in mosquitoes (*Diptera*: Culicidae). Systematic Zoology 33(3): 302–317

Rohlf FJ, Slice DS (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39(1): 40–59

Rumpunen K, Bartish IV (2002) Comparison of differentiation estimates based on morphometric and molecular data, exemplified by various leaf shape descriptors and RAPDs in the genus *Chaenomeles* (Rosaceae). Taxon 51(1): 69–82

Sampson PD, Bookstein FL, Sheehan FH, Boston EL (1996) Eigenshape analysis of left ventricular outlines from contrast ventriculograms. In Marcus LF, Corti M, Loy A, Naylor GJP, Slice D (Eds.) Advances in morphometrics, Plenum Press, New York, 211–234 pp

Sato T (1986) Life history characteristics of Polystichum tripteron with special reference to its leaf venation. The Botanical Magazine, Tokyo 99: 361–377

Sato T, Tsuyuzaki S (1988) Quantitative comparison of foliage development among Dryopteris monticola, D. tokyoensis, and a putative hybrid, D. kominatoensis in northern Japan. The Botanical Magazine, Tokyo 101: 267–280

Schneider H, Smith AR, Cranfill R, Hildebrand TE, Haufler CH, Ranker TA (2004) Unraveling the phylogeny of polygrammoid ferns (Polypodiaceae and Grammitidaceae): exploring aspects of the diversification of epiphytic plants. Molecular Phylogenetics and Evolution 31: 1041–1063

Schneller JJ (1990a) Azollaceae. In Kramer KU, Green PS (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin.

Schneller JJ (1990b) Salviniaceae. In Kramer KU, Green PS (Eds.) Volume I: Pteridophytes and gymnosperms. The families and genera of vascular plants, Kubitzki K (Ed.) Springer-Verglag, Berlin.

Schweitzer PN, Lohmann GP (1990) Life-history and the evolution of ontogeny in the ostracode genus *Cyprideis*. Paleobiology 16(2): 107–125

Sherwood RJ, Hlusko LJ, Duren DL, Emch VC, Walker A (2005) Mandibular symphysis of large-bodied homonoids. Human Biology 77(6): 735–759

Sille NP, Collinson ME, Kucera M, Hooker JJ (2004) Evolution within the charophyte genus *Harrisichara*, late Paleogene, southern England; environmental and biostratigraphic implications. Palaeogeography Palaeoclimatology Palaeoecology 208(1–2): 153–173

Sille NP, Collinson ME, Kucera M, Hooker JJ (2006) Morphological evolution of *Stratiotes* through the paleogene in England: an example of microevolution in flowering plants. PALAIOS 21(3): 272–288

Stein WE, Boyer JS (2006) Evolution of land plant architecture: beyond the telome theory. Paleobiology 32(3): 450–482

Teusch KP, Guralnick RP (2003) Environmentally driven variation in ancient populations of turritellids: evaluating the causal link. Paleobiology 29(2): 163–180

Torres MAJ, Demayo CG, Siar SV (2008) Elliptic Fourier analysis of leaf outline differences between and among sixteen species of Hoya. The Phillipine Agricultural Scientist 91(1): 18–28

Tryon RM (1964) Evolution of the leaf in living ferns. Memoirs of the Torrey Botanical Club 21: 73–82

Ubukata T (2003) A morphometric study on morphological plasticity of shell form in crevice-dwelling *Pterioida* (Bivalvia). Biological Journal of the Linnean Society 79(2): 285–297

Usukura M, Imaichi R, Kato M (1994) Leaf morphology of a facultative rheophyte, Farfugium japonicum var. Iuchuense (Compositae). Journal of Plant Research 107: 263–267

Wang Z, Chi Z, Feng D (2003) Shape based leaf image retrieval. Vision, Image and Signal Processing, IEEE Proceedings 150(1): 34–43

Weight C, Parnham D, Waites R (2008) LeafAnalyser: a computational method for rapid and large-scale analyses of leaf shape. The Plant Journal 53(3): 578–586

West JG, Noble IR (1984) Analyses of digitised leaf images of the *Dodonaea viscosa* complex in Australia. Taxon 33(4): 595–613

Whaley WH, Whaley CY (1942) A developmental analysis of inherited leaf patterns in *Tropaeolum*. American Journal of Botany 29(3): 195–200

White RJ, Prentice HC, Verwijst T (1988) Automated image acquisition and morphometric description. Canadian Journal of Botany 66: 450–459

Wiemann MC, Manchester SR, Dilcher DL, Hinjosa LF, Wheeler EA (1998) Estimation of temperature and precipitation from morphological characters of dicotyledonous leaves. American Journal of Botany 85: 1796–1802

Wilf P (1997) When are leaves good thermometers? A new case for Leaf Margin Analysis. Paleobiology 23: 373–390

Yahiaoui I, Herve N, Boujemaa N (2006) Shape-based image retrieval in botanical collections. In Lecture Notes in Computer Science 4261. Springer, Berlin, 357–364 pp

Young JP, Dickinson TA, Dengler NG (1995) A morphometric analysis of heterophyllous leaf development in *Ranunclus flabellaris*. International Journal of Plant Sciences 156(5): 590–602

Zahn CT, Roskies RZ (1972) Fourier descriptors for plane closed curves. IEEE Transactions on Computers C-21: 269–281

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) Geometric morphometrics for biologists: a primer. Elsevier Academic Press, New York, 443 p

Zurakowski KA, Gifford EM (1988) Quantitative studies of pinnule development in the ferns *Adiantum raddianum* and *Cheilanthes viridis*. American Journal of Botany 75(10): 1559–1570

Chapter 4
Discriminating Groups of Organisms

Richard E. Strauss

Idea and Aims

A common problem in morphometric studies is to determine whether, and in what ways, two or more previously established groups of organisms differ. Discrimination of predefined groups is a very different problem than trying to characterize the patterns of morphological variation among individuals, and so the kinds of morphometric tools used for these two kinds of questions differ. In this paper I review the basic procedures used for discriminating groups of organisms based on morphological characteristics – measures of size and shape. A critical reading of morphometric discrimination studies of various kinds of organisms in recent years suggests that a review of procedures is warranted, particularly with regard to the kinds of assumptions being made. I will discuss the main concepts and methods used in problems of discrimination, first using conventional morphometric characters (measured distances between putatively homologous landmarks), and then using landmarks directly with geometric morphometric approaches.

Introduction

Suppose that we have several sets of organisms representing two or more known groups. Individuals from the groups must be recognizable on the basis of extrinsic criteria. For example, if the groups represent females and males of some species of fish, then we might identify individuals using pigmentation patterns or other kinds of sexual sex characteristics or, lacking those, by examination of the gonads. The key idea is that we must be able to unambiguously assign individuals to previously recognized groups. We still might wish to know a number of things about them. Can we discriminate the groups based on morphometric traits? If so, how well? How different are the groups? Are the groups "significantly" different in morphology? How

R.E. Strauss (✉)
Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131, USA
e-mail: Rich.Strauss@ttu.edu

do we assess such significance in the presence of correlations among the morphometric characters? Which characters are the most important in discriminating the groups? Can group membership be predicted for "unknown" individuals? If so, how reliable are the predictions?

These questions can be answered (or at least approached) using three related kinds of methods: discriminant analysis (also called discriminant function analysis or canonical variate analysis), Mahalanobis distance, and multivariate analysis of variance. Discriminant analysis (DA) is used to estimate the linear combinations of characters that best discriminate the groups. Mahalanobis distance (D^2) estimates the distances between a pair of groups within the multivariate character space, in the presence of correlations among variables. And multivariate analysis of variance (MANOVA) determines whether the samples differ non-randomly (that is, significantly). It's interesting that the three kinds of methods were developed independently by three mathematicians: Fisher (DA) in England, Hotelling (MANOVA) in the United States, and Mahalanobis (D^2) in India. Due to differences in notation, underlying similarities between the methods were not noticed for some 20 years, but they now have a common algebraic formulation.

Conventional Morphometrics

Kinds of Data

Traditionally, before the onset of geometric morphometrics, morphometric studies were done using distances measured directly on specimens, often with calipers or microscopes, often in combination with meristic counts, angles, and other kinds of quantitative characters. Bookstein (Bookstein 1978; Bookstein et al. 1985; Strauss and Bookstein 1982) was the first to systematically stress the distinction between distances and other kinds of data, and the need to measure distances between comparable anatomical landmarks rather than arbitrarily on the form.

In the last decade or so, the use of digitizing equipment to record the positions of landmarks has become commonplace, and distances on specimens are usually calculated as Euclidean distances between landmarks. But directly measured distances continue to be used, sometimes mixed with other kinds of data.

For the following discussions I will assume that the variables (characters) consist entirely of distances measured between landmarks. Such distances are usually logarithmically transformed prior to analysis to improve their statistical properties and to characterize allometric relationships (Bookstein et al. 1985; Bryant 1986; Jungers et al. 1995; Keene 1995; Strauss 1993). However, use of log-transformations remains a somewhat controversial topic, and I won't pursue it here.

Principal Component Analysis

It's not uncommon for researchers to use principal component analysis (PCA) to attempt to discriminate groups of individuals. However, PCA is inherently a

single-group procedure and is not guaranteed to find group differences even if they exist. PCA is used to redistribute the total variance among a set of data points onto a set of mutually orthogonal axes (i.e., at right angles to one another) that merely redescribe the patterns of variation among the data. The new axes are the principal components, which are statistically independent of one another and so can be examined one at a time. The data points can be projected onto the axes (at right angles) to provide numerical scores of individuals on the components (Fig. 4.1). The principal components are calculated such that the variance of scores of individuals on the first axis (PC1) is as great as possible, so that PC1 can be said to account for the maximum variance in the data. Because the second component is by definition at right angles to the first, the scores of individuals on PC2 are uncorrelated with
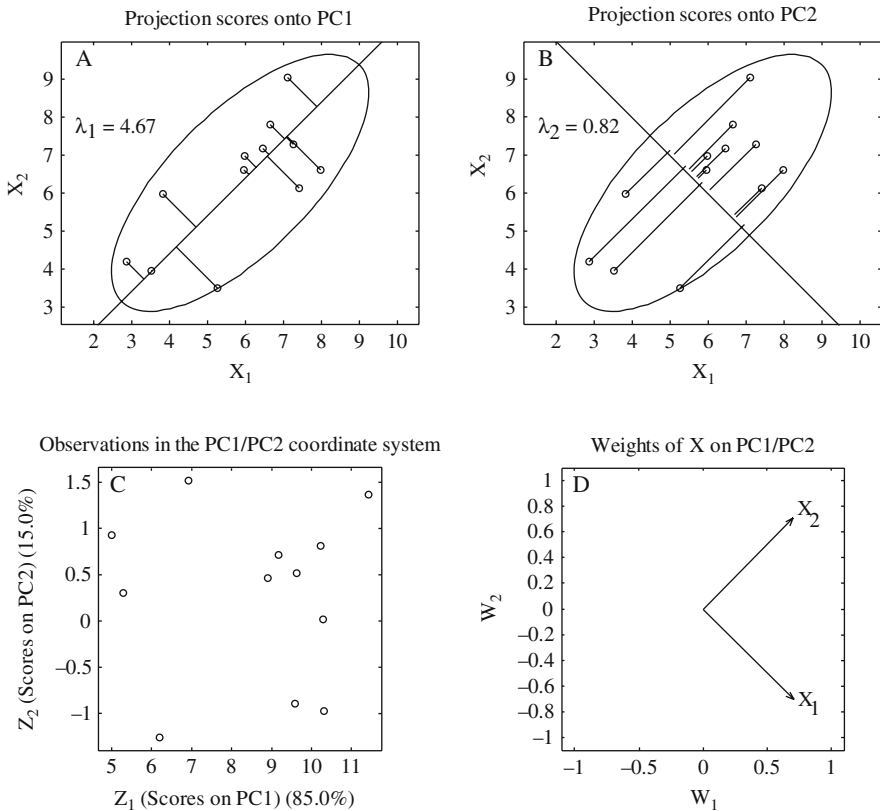

Fig. 4.1 Example of a principal component analysis for a scatter of data points for two variables. (**a**) The data points as projected onto PC1 to give scores on PC1. The ellipse is a 95% confidence interval on the data. The value λ_1 is the first eigenvalue, the variance of the scores on PC1. (**b**) The same data points as projected onto PC2. The value λ_2 is the second eigenvalue, the variance of the scores on PC2. (**c**) The data points plotted as scores in the space of components PC1 and PC2. (**d**) Projection of the axes for the two variables as unit vectors onto the space of components PC1 and PC2. These vectors indicate the maximum direction of variation in the corresponding variables in the PC1/PC2 space of **Panel C**

those on PC1. PC2 is the axis, orthogonal to PC1, on which the variance of scores is as great as possible. PC3 is the axis, mutually orthogonal to both PC1 and PC2, on which the variance of scores is as great as possible. And so on. PCA is usually used as a dimension-reduction procedure, because a scatterplot of points on the first two or three components may characterize most of the variation among the data point.

This procedure is a simple description of an eigenanalysis: the principal components are eigenvectors, and the variance of projection scores onto each component is the corresponding eigenvalue (Fig. 4.1). In practice, all components are calculated as a set rather than sequentially. The procedure can be viewed geometrically as a translation and solid rotation of the coordinate system. The origin of the coordinate system is moved (translated) to the center of the cloud of points, and then the coordinate axes are rotated as a set, at right angles to one another, so as to maximize the variance components. The data points maintain their original configuration, while the coordinate system moves around them. Thus the number of principal component axes is equal to the number of variables. The principal components are specified by sets of coefficients (weights, one per variable), the weights being computed so as to compensate for redundancy of information due to intercorrelations between variables. A principal component score for an individual is essentially a weighted average of the variables. The coefficients can be rescaled as vector correlations (Fig. 4.1), which are often more informative. The coefficients allow interpretation of the contributions of individual variables to variation in projection scores on the principal components. See Jolicoeur and Mosimann (1960) and Smith (1973) for early and very intuitive descriptions of the use of PCA in morphometric analyses.

The procedure inherently assumes that the data represent a single homogeneous sample from a population, although such structure isn't necessary to calculate the principal-component solution. (However, the assumption that the data were sampled from a multivariate-normally distributed population is necessary for classical tests of the significance of eigenvalues or eigenvectors.) Even if multiple groups are present in the data, the procedure does not take group structure into consideration. PCA maximizes variance on the components, regardless of its source. If the among-group variation is greater than the within-group variation, the PCA scatterplots might depict group differences. However, PCA is not guaranteed to discriminate groups. If group differences fail to show up on a scatterplot, it does not follow that group differences don't exist in the data.

Multiple-group modifications of PCA such as common principal components (CPC) have been developed (Flury 1988; Thorpe 1988), but these are generally not for purposes of discrimination. Rather, such methods assume that the same principal components exist in multiple groups (possibly with different eigenvalues) and allow estimation of the common components. Multiple-group methods are useful, for example, for adjusting morphometric data for variation in body size or other sources of extraneous variation prior to discrimination (Burnaby 1966; Humphries et al. 1981; Klingenberg et al. 1996).

Discriminant Analysis

In contrast to principal components analysis, discriminant analysis is explicitly a multiple-group procedure, and assumes that the groups are known (correctly) before analysis on the basis of extrinsic criteria and that all individuals are members of one (and only one) of the known groups. The terminology of discriminant analysis can be somewhat confusing. Fisher (1936) originally developed the "linear discriminant" for two groups. This was later generalized to the case of three or more groups independently by Bartlett, Hotelling, Mahalanobis, Rao and others to solve several related problems that are relevant to morphometric studies: the discrimination groups of similar organisms, the description of the morphological differences among groups, the measurement of overall difference between groups, and the allocation of "unknown" individuals to known groups. The allocation of unknown individuals is generally called classification, though this term is often used in a different way by systematic biologists, which by itself can cause confusion. The discrimination problem for three or more groups came to be known as "canonical variate analysis" ("canonical" in the sense of providing rules for classification), although this phrase has also been used synonymously with a related statistical procedure usually known as canonical correlation analysis. The tendency in recent years is to use "discriminant analysis" or "discriminant function analysis" for discrimination of any number of groups, although the term "canonical variate analysis" is still widely used.

Discriminant analysis (DA or DFA) optimizes discrimination between groups by one or more axes, the discriminant functions (DFs). These are mathematical functions in the sense that the projection scores of data points on the axes are linear combinations of the variables, as in PCA. Like PCA, DA is a form of eigenanalysis, except that in this case the axes are eigenvectors of the among-group covariance matrix rather than the total covariance matrix. For k groups, DA finds the k–1 discriminant axes that maximally separate the k groups (one axis for two groups, two for three groups, etc.). Like PCs, DFs have corresponding eigenvalues that specify the amount of among-group variance (rather than total variance) accounted for by the scores on each DF. Also like PCs, discriminant axes are linear combinations of the variables and are specified by sets of coefficients, or weights, that allow interpretation of contributions of individual variables. See Albrecht (1980, 1992) and Campbell and Atchley (1981) for geometric interpretations of discriminant analysis.

The first discriminant axis has a convenient interpretation in terms of analysis of variance of the projection scores (Fig. 4.2). Rather than being the axis that maximizes the total variance among scores, as in PCA (Fig. 4.1), the discriminant axis is positioned so as to maximize the total variance among groups relative to that within groups, which is the quantity measured by the ANOVA F-statistic. The projection scores on DF1 give an F-statistic value greater than that of any other possible axis. The same is true for three or more groups (Fig. 4.3). The DF1 axis is positioned so as to maximize the dispersion of scores of groups along it. The dispersion giving the
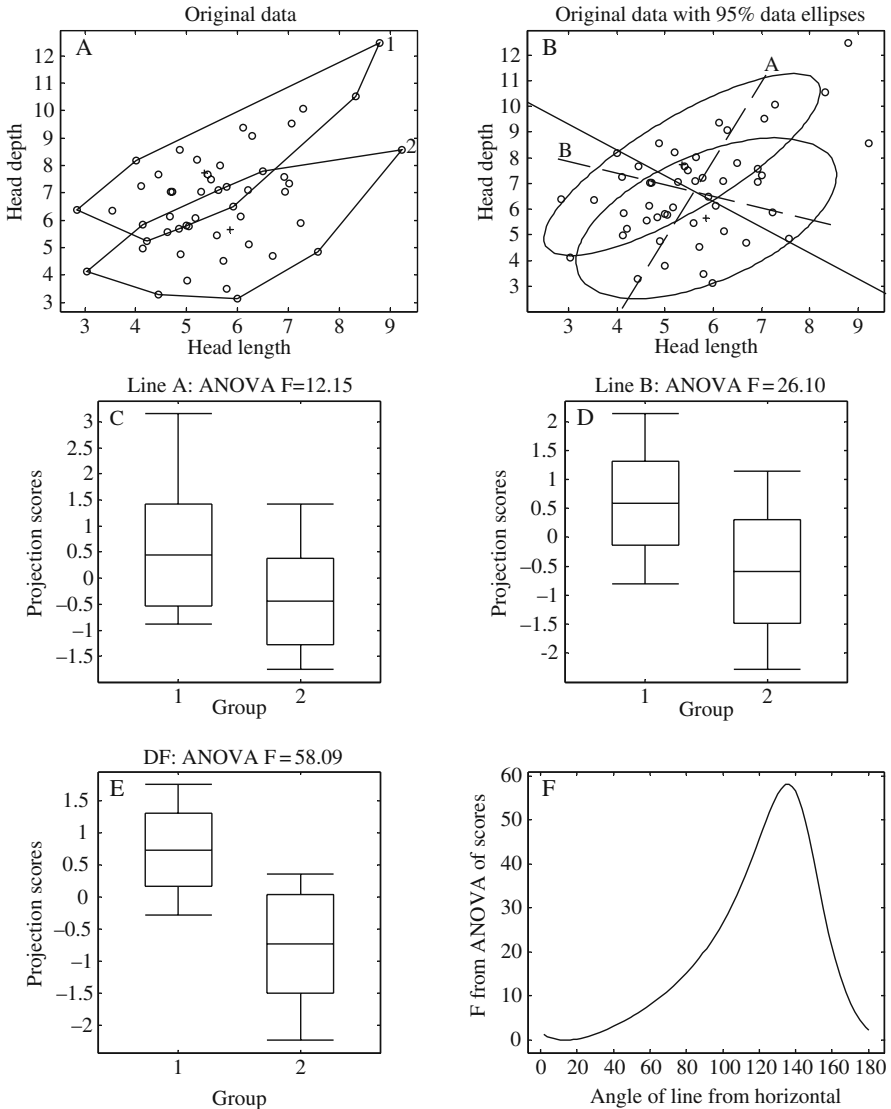
Fig. 4.2 Example of a discriminant analysis for samples of two species of *Poecilia*, in terms of two variables: head length and head width, both in mm. (**a**) Original data, with convex hulls indicating dispersion of data points for the two groups. (**b**) Data and 95% confidence intervals for the two groups. *Dotted lines* A and B are arbitrarily chosen axes; the *solid line* is the discriminant axis for the two groups. (**c**) Box plots for the two groups of projection scores onto *dotted line* A, and corresponding ANOVA F-statistic. (**d**) Box plots of projection scores onto *dotted line* B, and corresponding ANOVA F-statistic. (**e**) Box plots of projection scores onto the discriminant axis, and corresponding ANOVA F-statistic. (**f**) F-statistic from ANOVAs of projection scores onto all possible axes, as a function of angle (in degrees) from the horizontal (head-length axis) of **Panel B**. The discriminant axis is that having the maximum ANOVA F-statistic value
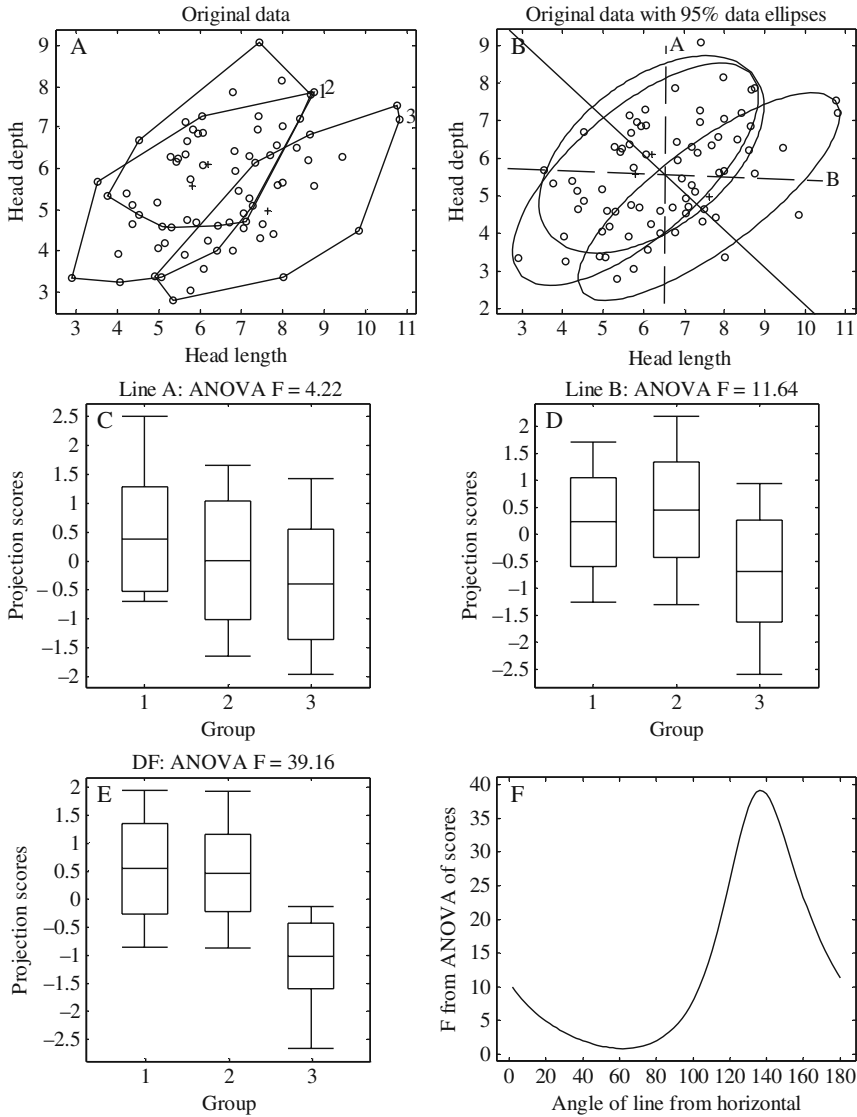
Fig. 4.3 Example of a discriminant analysis for samples of three species of *Poecilia*, in terms of two variables: head length and head width, both in mm. (**a**) Original data, with convex hulls indicating dispersion of data points for the three groups. (**b**) Data and 95% confidence intervals for the two groups. *Dotted lines* A and B are arbitrarily chosen axes; the *solid line* is the discriminant axis for the two groups. (**c**) Box plots for the three groups of projection scores onto *dotted line* A, and corresponding ANOVA F-statistic. (**d**) Box plots of projection scores onto *dotted line* B, and corresponding ANOVA F-statistic. (**e**) Box plots of projection scores onto the discriminant axis, and corresponding ANOVA F-statistic. (**f**) F-statistic from ANOVAs of projection scores onto all possible axes, as a function of angle (in degrees) from the horizontal (head-length axis) of **Panel B**. The discriminant axis is that having the maximum ANOVA F-statistic value

maximum F-statistic might distinguish one group from the others (as in Fig. 4.3e), or might separate all groups by a small amount; the particular pattern depends on the structure of the data.

As with PCA, a unique set of discriminant axes can be calculated for any set of data if the sample sizes are sufficiently large. However, inferences about the populations from which the data were sampled are reasonable only if the populations are assumed to be multivariate-normally distributed with equal covariance matrices (the multivariate extensions of the normality and homoscedasticity assumptions of ANOVA). In particular, discrimination of samples will be optimal with respect to their populations only if this distributional assumption is true. Because a topological cross-section through a multivariate normal distribution is an ellipse, confidence ellipses on the sample data are often depicted on scatterplots to visually assess this underlying assumption (Owen and Chmielewski 1985; Figs. 4.2b and 4.3b). If the assumption about population distributions is true, then the sample ellipses will be approximately of the same size and shape because they will differ only randomly (i.e., they will be homogeneous). Bootstrap and other randomization methods can give reliable confidence intervals on estimates of discriminant functions and related statistics even if the distributional assumption is violated (Dalgleish 1994; Ringrose 1996; Von Zuben et al. 1998; Weihs 1995).

The minimum sample sizes required for a discriminant analysis can sometimes be limiting, particularly if there are many variables relative to the number of specimens, as is often the case in morphometric studies. In the same way that an analysis of variance of a single variable is based on the among-group variance relative to the pooled (averaged) within-group variance, in a DA the eigenvectors and eigenvalues are derived from the among-group covariance matrix relative to the pooled within-group covariance matrix, which is the averaged covariance matrix across all groups. Using the pooled matrix is reasonable if the separate matrices differ only randomly, as assumed. But if the separate matrices are quite different, they can average out to a "circular" rather than elliptical distribution, for which the net correlations are approximately zero. In this case the DA results would not differ much from those of a PCA.

The minimum sample size requirement for a DA relates to the fact that the pooled within-group matrix must be inverted (because it's "in the denominator", so to speak), and inversion can't be done unless the degrees of freedom of the within-group matrix be greater than the number of variables. The within-group degrees of freedom is typically N-p-1, where N is the total sample size and p is the number of variables. However, this is the minimum requirement for a solution to be found. The number of specimens should be much larger than the number of variables for a stable solution – one that wouldn't change very much if a new set of samples from the same populations were taken. A typical rule of thumb is that the number of specimens should be at least five or so times the number of variables. However, the minimally reasonable sample size depends on how distinctive the groups are (because subtle differences require more statistical power to detect). In addition, it requires larger sample sizes to determine the nature of the differences among groups than just to demonstrate that the difference is significant.
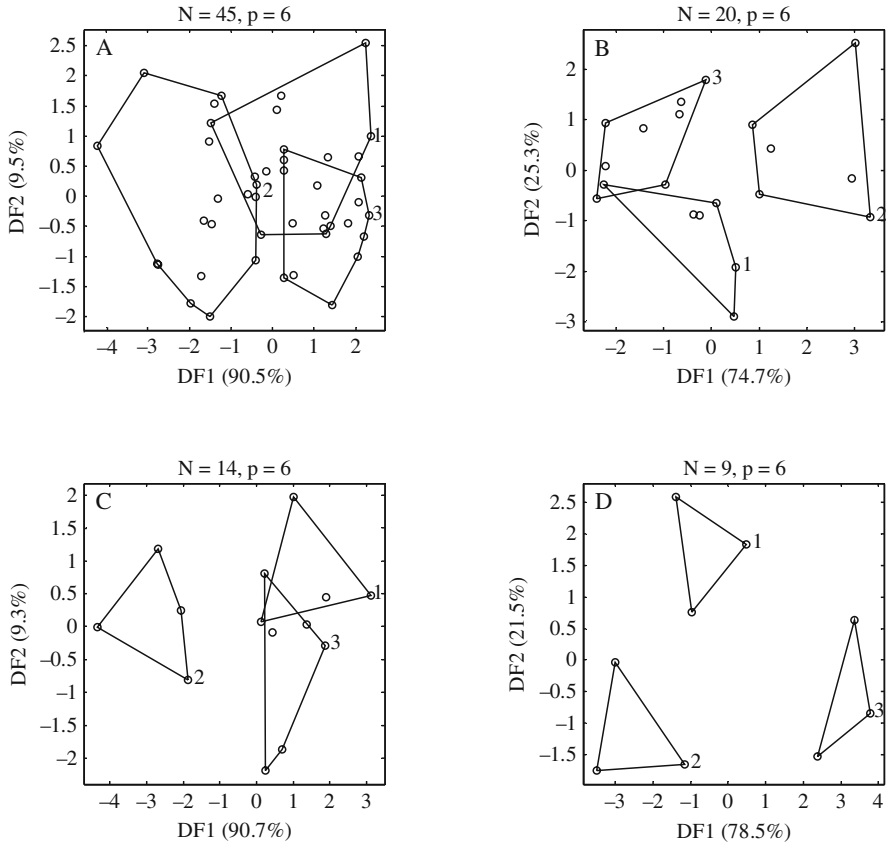
Fig. 4.4 Example of the effect of sample size on the apparent discrimination among three groups. (**a**) Scatterplot of scores on the two discriminant axes for 45 specimens and 6 variables. (**b**) Scatterplot of scores for a random 20 of 45 specimens. (**c**) Scatterplot of scores for a random 14 of 45 specimens. (**d**) Scatterplot of scores for a random 9 of 45 specimens

Because of this matrix-inversion problem, the degree of discrimination among groups can become artificially inflated for small sample sizes (relative to the number of variables) (Fig. 4.4). Scatterplots on discriminant axes can suggest that groups are highly distinctive even though the group means might actually differ by little more than random variation. Because of this, discriminant scatterplots must be interpreted with caution, and never without supporting statistics (described below).

Another factor that enters into the minimum-sample-size issue is variation in the number of specimens per group. When the covariance matrices for the separate groups are pooled, the result is a weighted average covariance matrix, weighted by sample size per group. This makes sense because the precision of any statistical estimate increases as sample size increases, and so a covariance matrix for a large sample is a more reliable estimate of the "real" covariance matrix. Because

variances and covariances can be estimated for as few as three specimens, very small groups can in principle be included in a discriminant analysis. In practice, however, it is often beneficial to omit groups having sample sizes of less than five or so.

Some recently developed methods for performing discriminant analysis with relatively small sample sizes (e.g. Anderson and Robinson 2003; Howland and Park 2004; Ye et al. 2004) seem promising, but none have yet been applied to morphometric data.

Size-Free Discriminant Analysis

In systematics it has long been considered desirable to be able to discriminate among groups of organisms (populations, species, etc.) on the basis of "size-free" or size-invariant shape measures (dos Reis et al. 1990; Humphries et al. 1981). This is particularly important when the organisms display indeterminant growth, in which case discrimination among taxa might represent merely a sampling artifact if different samples comprise different proportions of age classes. Discrimination among samples in which variation in size cannot be easily controlled may lead to spurious results, since the size-frequency distribution of different taxa will be a function of the ontogenetic development of individuals present in different samples. In this case one way of correcting the problem would be to statistically "correct" or adjust for the effect of size present within samples of each group. However, a number of different definitions of "size-free" shape have been applied. The terms shape and size have been used in various and sometimes conflicting ways (Bookstein 1989a).

In size adjustment the effects of size variation are to be partitioned or removed from the data, usually by some form of regression, and residuals are subsequently used as size-independent shape variables (Jolicoeur et al. 1984; Jungers et al. 1995). In distance-based morphometrics, the most common methods for size adjustment have involved bivariate regression (Albrecht et al. 1993; Schulte-Hostedde et al. 2005; Thorpe 1983) multiple-group principal components (Pimentel 1979; Thorpe and Leamy 1983), sheared principal components (Bookstein et al. 1985; Humphries et al. 1981; Rohlf and Bookstein 1987), and Burnaby's procedure (Burnaby 1966; Gower 1976; Rohlf and Bookstein 1987). Although many different methods have been proposed, there has been little agreement on which method should be used. This issue is important because different size-adjustment methods often yield slightly different results.

In the case of size-adjustment for multiple taxa, the issue arises as to whether and how group structure (e.g., presence of multiple species) should be taken into consideration (Klingenberg and Froese 1991) – whether the correction should be separately by group or should be based on the pooled within-group regression. The latter implicitly assumes that all within-group covariance matrices are identical, although this assumption can be relaxed with use of common principal components (Airoldi and Flury 1988; Bartoletti et al. 1999; Klingenberg et al. 1996).

Mahalanobis Distances

Whereas discriminant analysis scores can provide a visualization of group separation, Mahalanobis distances (D^2) measure the distances between group centroids on a scale that is adjusted to the (pooled) within-group variance in the direction of the group difference. (D, the square root of D^2, measures the distance between group centroids adjusted by the standard deviation rather than the variance.) In Fig. 4.5, for example, the Euclidean (straight-line) distance from centroid A to centroid B is that same as that from A to C. However, the Mahalanobis distances are quite different because the distance from A to B is measured "with the grain" while that from A to C is measured "across the grain". In terms of variation, the relative distance from A to C is much greater than that from A to B.

This is often said to be analogous to using an F-statistic to measure the difference between two group means, although that is not quite correct – an F-statistic increases as sample size increases, whereas a Mahalanobis distance approaches its "true" value with increasing sample size. The Mahalanobis distance is essentially a distance in a geometric space in which the variables are uncorrelated and equally scaled. It also possesses all of the characteristics that a measure must have to be a metric: the distance between two identical points must be zero, the distance between two non-identical points must be greater than zero, the distance from A to B must be the same as that from B to A (symmetry), and the pairwise distances among three points must satisfy the triangle inequality. For morphometric data, such a measure of group separation is more informative than the simple Euclidean distance between groups.
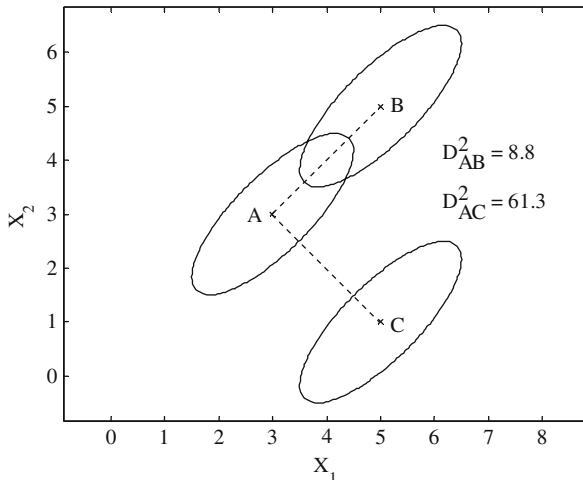

Fig. 4.5 Mahalanobis distances between centroids of groups. Variation within groups is indicated by 95% confidence ellipses for the data. Euclidean distances between centroids of A and B and of A and C are both 2.83. Corresponding Mahalanobis distances are indicated on plot

Mahalanobis distances can also be measured between a point and a group centroid or between two points. In both cases the distance is relative to the covariance matrix of the group.

Confidence intervals for Mahalanobis distances can be estimated by comparison to a theoretical F distribution if the distribution of the group(s) is assumed to be multivariate normal (Reiser 2001). More robust confidence intervals for real biological data can be estimated by bootstrapping the data within-group (Edgington 1995; Manly 1997; Wilcox 2005).

MANOVA

Analysis of variance (ANOVA) is the univariate case of the more general multivariate analysis of variance (MANOVA). Instead of a "univariate F" statistic measuring the heterogeneity among a set of means with respect to the pooled within-group variance, the resulting "multivariate F" measures the heterogeneity among a set of multivariate centroids with respect to the pooled within-group covariance matrix. The covariance matrix accounts for the observed correlations among variables. As with ANOVA, the samples can be cross-classified with respect to two or more factors, or can be structured with respect to other kinds of sampling designs (Gower and Krzanowski 1999).

In practice the actual test statistic calculated is Wilks' lambda, which is related to the computations involved in discriminant functions and Mahalanobis distances. It is a direct measure of the proportion of total variance in the variables that is not accounted for by the grouping of specimens. If Wilks' lambda is small, then a large proportion of the total variance is accounted for by the grouping, which in turns suggests that the groups have different mean values for one or more of the variables. Because the sampling distribution of Wilks' lambda is rather difficult to evaluate, lambda is usually transformed approximately to an F statistic. There are a number of alternative statistics that are similar in purpose to Wilks' lambda but that have somewhat different statistical properties, such as Pillai's trace and Roy's greatest root. These are often reported by statistical software, but in general are not widely used (Everitt and Dunn 2001).

Under the null hypothesis that all groups have been sampled randomly from the same population, and therefore differ only randomly in all of their statistical properties, the F statistic can be used to estimate a "P-value", the probability of sampling the observed amount of heterogeneity among centroids if the null hypothesis is true. The P-value is accurate only if the population from which the groups have been sample is multivariate-normal in distribution. If the null hypothesis is true, then the covariance matrices for all groups will differ only randomly (i.e., they will be homogeneous), and thus can be pooled for the test. If the within-group covariance matrices differ significantly, then the pooled covariance matrix may be biased, as will the P-value. As with statistical tests in general, violated assumptions will often (but not necessarily) lead to P-values that are too small, and thus will lead to the rejection of the null hypothesis too often.

Since claiming significant differences when they don't exist is counterproductive in science, the dependence of MANOVA on such stingent assumptions is a problem. This can be circumvented to some degree by using randomization procedures (e.g., random permutation) to estimate the null sampling distribution of the test statistic rather than theoretical distributions (such as the F distribution) (Anderson 2001). Such "non-parametric" tests, although not assumption-free, tend to be much more robust to statistical assumptions than are conventional statistical hypothesis tests.

It is often assumed that a series of separate ANOVAs, one per variable, is equivalent to a MANOVA. However, this is not the case, for several reasons (Willig and Owen 1987). First, if the variables are correlated, then the separate ANOVAs are not statistically independent. For example, if the ANOVA for one variable is statistically significant, then the ANOVAs for variables correlated with it will also tend to be significant. Thus the results from the ANOVAs will be redundant to an unknown extent and difficult to integrate. Second, the overall ("family-wise") Type I error rate become artificially high as the number of statistical tests increases, so that the probability of obtaining a significant results due to chance increases (the "multiple-comparisons" problem; Hochberg and Tamhane 1987).

If the overall MANOVA is statistically significant, then separate ANOVAs can be done to assess which of the variables has contributed to the group differences. But the multiple-comparisons issues remain, and subsequent statistical testing must be done carefully.

Classification

A procedure closely related to discriminant functions and Mahalanobis distances is that of classifying "unknown" specimens to known, predefined groups. (Note that this use of "classification" is related to, but different from, the common use of the term in systematics.) A strong assumption of any classification procedure is that the individual being classified is actually a member of one of the groups included in the analysis. If this assumption is ignored or wrong, then any estimated probabilities of group membership may be misleading (Albrecht 1992).

There are two basic approaches to classifying unknowns with morphometric data. The first, and most conventional, is based in principle on means: calculate the Mahalanobis distance from the unknown to the centroid of each group, and assign it to the closest group (Hand 1981). Because Mahalanobis distances are based on pooled covariance matrices, correct assignments depend on the assumptions of homogeneous covariance matrices and, to a lesser degree, of multivariate normality. This approach can be viewed as subdividing the data space into mutually exclusive "decision spaces", one for each predefined group, and classifying each unknown according to the decision space in which it lies. Each Mahalanobis distance has an associated chi-square probability, which can be used to estimate probabilities of group membership (or their complements, probabilities of misclassification; Williams 1982). More robust estimates of classification probabilities can be

approximated by bootstrapping the "known" specimens within-group (Davison and Hinkley 1996; Fu et al. 2005; Higgins and Strauss 2004).

The second approach is to view the data space in terms of mixtures of multivariate-normal distributions, one for each predefined group. Such methods tend to be much more sensitive to deviations from the assumptions of multivariate normality and homogeneous covariance matrices, but can better accommodate differences in sample size among groups (White and Ruttenberg 2007).

Cross Validation

Cross-validation is a widely used resampling technique that is often used for the assessment of statistical models (Stone 1974). Like other randomization methods such as the bootstrap and jackknife, it is almost distribution-free in the sense that it evaluates the performance of a statistical procedure given the actual structure of the data. It is necessary because whenever predictions from a statistical model are evaluated with the same data used to estimate the model, the fit is "too good"; this is known as over-fitting. When new data are used, the model almost always performs worse than expected. In the case of discriminant analysis and related methods, overfitting comes into play both in the assessment of group differences (discriminant-score plots and MANOVA) and in estimates of probabilities of group membership.

The basic idea behind cross-validation is simply to use a portion of the data (the "training" or "calibration" set) to fit the model and estimate parameters, and use the remaining data (the "test" set) to evaluate the performance of the model. For classification problems, for example, the group identities of all specimens are known in advance, and so they can be used to check whether the predicted identities are correct. This is typically done in a "leave-one-out" manner: one specimen is set aside and all N-1 others are used to estimate Mahalanobis distances. The omitted specimen is then treated as an unknown and its group membership is predicted. The procedure is repeated for all specimens, sequentially leaving each one out of the analysis and estimating distances from the others, then predicting the group membership of the omitted specimen. The overall proportions of correct predictions are unbiased estimates of the probabilities of correct classification, given the actual structure of the data. Cross-validation methods are particularly appropriate for small samples (Fu et al. 2005).

Related Methods

The most commonly used alternative to discriminant analysis is logistic regression, which usually involves fewer violations of assumptions, is robust, handles discrete and categorical data as well as continuous variables, and has coefficients that are somewhat easier to interpret (Hosmer and Lemeshow 2000). However, discriminant analysis is preferable when its assumptions are reasonably met because it has consistently greater statistical power (Press and Wilson 1978).

Quadratic discriminant analysis (QDA) is closely related to linear discriminant analysis (LDA), except that there is no assumption that the covariance matrices of the groups are homogeneous (Meshbane and Morris 1995). When the covariance matrices are homogeneous, LDA is systematically better than QDA both at group separation and classification. When the covariance matrices vary significantly, QDA is usually better, but not always, especially for small samples (Flury et al. 1994; Marks and Dunn 1974). This is apparently due to the greater robustness of LDA to violation of assumptions. In any case, there have been few morphometric applications of quadratic discriminant analysis.

There are several different versions of nonlinear discriminant analysis, which finds nonlinear functions that best discriminate among known groups. Most nonlinear methods work by finding some linear transformation of the character space that produces optimal linear discriminant functions. Generalized discriminant analysis (Baudat and Anouar 2000) has become the most widely used method.

And finally, neural networks have been used successfully in both linear and nonlinear discrimination and classification problems (Baylac et al. 2003; Dobigny et al. 2002; Higgins and Strauss 2004; Kiang 2003; Raudys 2001; Ripley 1994).

Geometric Morphometrics

Whereas conventional morphometric studies utilize distances as variables, geometric morphometrics (Bookstein 1991; Dryden and Mardia 1998; Rohlf 1993) is based directly on the digitized x,y,(z)-coordinate positions of landmarks, points representing the spatial positions of putatively homologous structures in two or three dimensions. Bookstein (1991) has characterized the types of landmarks, their configurations, and limitations, and Adams (1999) has extended their utility.

Once landmark coordinates have been obtained for a set of forms, they must be standardized to be directly comparable. This is typically done using a generalized Procrustes analysis in two or three dimensions, in which the sum of squared distances between homologous landmarks of each form and a reference configuration is iteratively minimized by translations and rigid rotations of the landmark configurations (Goodall 1995; Gower 1975; Penin and Baylac 1995; Rohlf and Slice 1990).

Isometric size differences are eliminated by dividing the coordinates of each form by its centroid size, defined as the square root of the sum of the squared distances between the geometric center of the form and its landmarks (Bookstein 1991). The residual variation in landmark positions among forms (deviations from the reference form) are referred to as "Procrustes residuals" in the x and y (and possibly z) coordinate directions. The square root of the sum of the squared distances between corresponding landmarks of two aligned configurations is an approximation of Procrustes distance, which plays a central role in the theory of shape analysis (Small 1996). It is also the measure that binds together the collection of methods for the analysis of shape variation that comprises the "morphometric synthesis" (Bookstein 1996).

To characterize and visualize differences between pairs of reference forms, the aligned landmark coordinates are often fitted to an interpolation function such as a thin-plate spline (Bookstein 1989b; Rohlf and Slice 1990), which can be decomposed into global (affine) and local (nonaffine) components. The nonaffine component can be further decomposed into partial or relative warps, geometrically orthogonal (and thus independent) components that correspond to shape deformations at different scales.

However, for the purpose of discrimination among groups of forms, the Procrustes residuals can be used directly as variables for discriminant analysis, MANOVA, and classification, as described above. In this case the number of variables for two-dimensional forms is twice the number of landmarks, one set for the x coordinates and one set for the y coordinates. For three-dimensional forms the number of variables would be three times the number of landmarks.

The large number of variables relative to the number of specimens therefore presents even more of a problem in geometric morphometrics than it tends to do in conventional morphometrics. The usual procedure is to use the Procrustes residuals in a principal component analysis, and then use the projection scores on the first few components as derived variables (e.g., Depecker et al. 2006). Since these derived variables are uncorrelated across all observations, the covariance matrices have zeros in the off-diagonal positions, and Mahalanobis distances are equivalent to Euclidean distances.

Conclusion

The multivariate methods reviewed here remain a powerful set of tools for morphometric studies, and their importance in the field cannot be overemphasized. Although the widespread availability of computer software has permitted their use by biologists of varying levels of statistical background and sophistication, it remains true that it is the responsibility of individual researchers to understand the properties and underlying assumptions of the methods they use.

References

Adams DC (1999) Methods for shape analysis of landmark data from articulated structures. Evolutionary Ecology Research 1: 959–970

Airoldi JP, Flury BK (1988) An application of common principal component analysis to cranial morphometry of *Microtus californicus* and *Microtus ochrogaster* (Mammalia, Rodentia). Journal of Zoology (London) 216: 21–36

Albrecht GH (1980) Multivariate analysis and the study of form, with special reference to canonical variate analysis. American Zoologist 20: 679–693

Albrecht GH (1992) Assessing the affinities of fossils using canonical variates and generalized distances. Human Evolution 7: 49–69

Albrecht GH, Gelvin BR, Hartman SE (1993) Ratios as a size adjustment in morphometrics. American Journal of Physical Anthropology 91: 441–468

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32–46

Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. Australian and New Zealand Journal of Statistics 45: 301–318

Bartoletti S, Flury BD, Nel DG (1999) Allometric extension. Biometrics 55: 1210–1214

Baudat G, Anouar F (2000) Generalized discriminant analysis using a kernel approach. Neural Computation 12: 2385–2404

Baylac M, Villemant C, Simbolotti G (2003) Combining geometric morphometrics with pattern recognition for the investigation of species complexes. Biological Journal of the Linean Society 80: 89–98

Bookstein FL (1978) The Measurement of biological shape and shape change. Lecture Notes in Biomathematics 24. Springer-Verlag, New York

Bookstein FL (1989a) "Size and shape": a comment on semantics. Systematic Zoology 38: 173–180

Bookstein FL (1989b) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11: 567–585

Bookstein FL (1991) Morphometric tools for landmark data: geometry and biology, Cambridge University Press, Cambridge

Bookstein FL (1996) Combining the tools of geometric morphometrics. In Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE (Eds.) Advances in morphometrics, Plenum Press, New York, 131–151 pp

Bookstein FL, Chernoff B, Elder RL, Humphries JM, Smith GR, Strauss RE (1985) Morphometrics in evolutionary biology: the geometry of size and shape change. Academy of Natural Sciences, Philadelphia

Bryant EH (1986) On use of logarithms to accommodate scale. Systematic Zoology 35: 552–559

Burnaby TP (1966) Growth-invariant discriminant functions and generalized distances. Biometrics 22: 96–110

Campbell NA, Atchley WR (1981) The geometry of canonical variates analysis. Systematic Zoology 30: 268–280

Dalgleish LI (1994) Discriminant analysis: statistical inference using the jackknife and bootstrap procedures. Psychological Bulletin 116: 498–508

Davison AC, Hinkley DV (1996) Bootstrap methods and their application, Cambridge University Press, Cambridge

Depecker M, Berge C, Penin X, Renous S (2006) Geometric morphometrics of the shoulder girdle in extant turtles (Chelonii). Journal of Anatomy 208: 35–45

Dobigny G, Baylac M, Denys C (2002) Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (Rodentia, Gerbillinae). Biological Journal of the Linnean Society 77: 319–327

dos Reis SF, Pessoa LM, Strauss RE (1990) Application of size-free canonical discriminant analysis to studies of geographic differentiation. Revista Brasileira de Genética 13: 509–520

Dryden IL, Mardia KV (1998) Statistical shape analysis. John Wiley, New York

Edgington ES (1995) Randomization tests. Marcel Dekker, New York

Everitt BS, Dunn G (2001) Applied multivariate data analysis, Second Edition. Wiley & Sons, New York

Fisher RA (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics 7: 179–188

Flury B, Schmid MJ, Narayanan A (1994) Error rates in quadratic discrimination with constraints on the covariance matrices. Journal of Classification 11: 101–120

Flury BK (1988) Common principal components and related multivariate models, Wiley, New York

Fu WJ, Carroll RJ, Wang S (2005) Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 21: 1979–1986

Goodall CR (1995) Procrustes methods in the statistical analysis of shape revisited. In Mardia KV, Gill CA (Eds.) Current issues in statistical shape analysis, University of Leeds Press, Leeds, 18–33 pp

Gower JC (1975) Generalized procrustes analysis. Psychometrika 40: 33–51

Gower JC (1976) Growth-free canonical variates and generalized inverses. Bulletin of the Geological Institutions of the University of Uppsala 7: 1–10

Gower JC, Krzanowski WJ (1999) Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. Royal Statistical Society: Series C (Applied Statistics) 48: 505–519

Hand DJ (1981) Discrimination and classification. John Wiley, New York

Higgins CL, Strauss RE (2004) Discrimination and classification of foraging paths produced by search-tactic models. Behavioral Ecology 15: 248–254

Hochberg Y, Tamhane AC (1987) Multiple comparison procedures, Wiley & Sons, New York

Hosmer DW, Lemeshow S (2000) Applied logistic regression analysis, Second Edition, John Wiley & Sons, New York

Howland P, Park H (2004) Generalizing discriminant analysis using the generalized singular value decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 995–1006

Humphries JM, Bookstein FL, Chernoff B, Smith GR, Elder RL, Poss S (1981) Multivariate discrimination by shape in relation to size. Systematic Zoology 30: 291–308

Jolicoeur P, Mosimann JE (1960) Size and shape variation in the painted turtle: a principal component analysis. Growth 24: 399–354

Jolicoeur P, Pirlot P, Baron G, Stephan H (1984) Brain structure and correlation patterns in Insectivora, Chiroptera, and primates. Systematic Zoology 33: 14–29

Jungers WL, Falsetti AB, Wall CE (1995) Shape, relative size, and size-adjustments in morphometrics. Yearbook of Physical Anthropology 38: 137–161

Keene ON (1995) The log transformation is special. Statistics in Medicine 14: 811–819

Kiang MY (2003) A comparative assessment of classification methods. Decision Support Systems 35: 441–454

Klingenberg CP, Froese R (1991) A multivariate comparison of allometric growth patterns. Systematic Zoology 40: 410–419

Klingenberg CP, Neuenschwander BE, Flury BD (1996) Ontogeny and individual variation: analysis of patterned covariance matrices with common principal components. Systematic Biology 45: 135–150

Manly BFJ (1997) Randomization, bootstrap and Monte Carlo methods in biology. Chapman & Hall, London

Marks S, Dunn OJ (1974) Discriminant functions when covariance matrices are unequal. Journal of the American Statistical Association 69: 555–559

Meshbane A, Morris JD (1995) A method for selecting between linear and quadratic classification models in discriminant analysis. Journal of Experimental Education 63: 263–273

Owen JG, Chmielewski MA (1985) On canonical variates analysis and the construction of confidence ellipses in systematic studies. Systematic Zoology 34: 366–374

Penin X, Baylac M (1995) Analysis of skull shape changes in apes, using 3D Procrustes superimposition. In Mardia KV, Gill CA (Eds.) Current issues in statistical shape analysis, Leeds University Press, Leeds, England, 208–210 pp

Pimentel RA (1979) Morphometrics: the multivariate analysis of biological data. Kendall-Hunt, Dubuque

Press SJ, Wilson S (1978) Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73: 699–705

Raudys S (2001) Statistical and neural classifiers: an integrated approach. Springer, New York

Reiser B (2001) Confidence intervals for the Mahalanobis distance. Communications in Statistics – Simulation and. Computation 30: 37–45

Ringrose TJ (1996) Alternative confidence regions for canonical variate analysis. Biometrika 83: 575–587

Ripley BD (1994) Neural networks and related methods for classification. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 56: 409–456

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. In Marcus LF, Bello E, Garcia-Valdecasas A (Eds.) Contributions to morphometrics, Museo Nacional de Ciencias Naturales, Madrid, Spain, 131–159 pp

Rohlf FJ, Bookstein FL (1987) A comment on shearing as a method for "size correction". Systematic Zoology 36: 356–367

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superposition of landmarks. Systematic Zoology 39: 40–59

Schulte-Hostedde AI, Zinner B, Millar JS, Hickling GJ (2005) Restitution of mass-size residuals: validating body condition indices. Ecology 86: 155–163

Small CG (1996) The statistical theory of shape. Springer, New York

Smith GR (1973) Analysis of several hybrid cyprinid fishes from western North America. Copeia 1973: 395–410

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 36: 111–147

Strauss RE (1993) The study of allometry since Huxley. In Huxley JS (Ed.) Problems of relative growth, Johns Hopkins University Press, Baltimore, 47–75 pp

Strauss RE, Bookstein FL (1982) The truss: body form reconstruction in morphometrics. Systematic Zoology 31: 113–135

Thorpe RS (1983) A review of the numerical methods for recognising and analysing racial differentiation. In Felsenstein J (Ed.) Numerical taxonomy, Springer-Verlag, Berlin, 404–423 pp

Thorpe RS (1988) Multiple group principal components analysis and population differentiation. Journal of Zoology (London) 216: 37–40

Thorpe RS, Leamy L (1983) Morphometric studies in inbred and hybrid housemice (*Mus* sp.): multivariate analysis of size and shape. Journal of Zoology (London) 199: 421–432

Von Zuben FJ, Duarte LC, Stangenhaus G, Pessoa LM, dos Reis SF (1998) Bootstrap confidence regions for canonical variates: application to studies of evolutionary differentiation. Biometrical Journal 40: 327–339

Weihs C (1995) Canonical discriminant analysis: comparison of resampling methods and convex-hull approximation. In Krzanowski WJ (Ed.) Recent advances in descriptive multivariate analysis, Clarendon Press, Oxford, UK, 34–50 pp

White JW, Ruttenberg BI (2007) Discriminant function analysis in marine ecology: some oversights and their solutions. Marine Ecology Progress Series 329: 301–305

Wilcox RR (2005) Introduction to robust estimation and hypothesis testing, Academic Press, New York

Williams BK (1982) A simple demonstration of the relationship between classification and canonical variates analysis. American Statistician 36: 363–365

Willig MR, Owen RD (1987) Univariate analyses of morphometric variation do not emulate the results of multivariate analyses. Systematic Zoology 36: 398–400

Ye J, Janardan R, Park CH, Park H (2004) An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 982–994

Chapter 5
Visual Analysis in Archaeology. An Artificial Intelligence Approach

Juan A. Barceló

Idea and Aims

Archaeology is a quintessentially "visual" discipline, because visual perception makes us aware of fundamental properties of objects and allows us to discover how objects were produced and used in the past. The approach I adopt here is to follow current computational theories of visual perception to ameliorate to way archaeology can deal with the analysis and explanation of the most usual visual marks: shape and texture. In any case, I am not interested in the mere mechanical procedure of extracting shape information among visual input, but in explaining why archaeological evidences have the shape they have.

Introduction

Archaeologists are interested in finding the social *cause* (production, use, distribution) of what they "see" at the archaeological site (or at the museum collection). By assuming, that what they perceive in the present is simply the material effects of human work made in the past, archaeologists try to understand "archaeological percepts" as material things that were products configured through human labor at the very beginning of their causal history (Barceló 2007).

The first we have to take into account when dealing with archaeology is that it is a quintessentially "visual" discipline. Among all features that describe archaeological evidences, some of them, the most important for the recognition and/or the discovery of the way the item was produced and or used in the past, have something to do with what we have been trained to "see". Tasks such as identifying the nature of the evidence, an artifact type, identifying decorative patterns or use wear in archaeological materials, recognizing archaeological structures in a satellite or aerial image,

J.A. Barceló (✉)

Departament de Prehistòria, Facultat de Lletres, Universitat Autònoma de Barcelona, Campus Bellaterra, 08193 Cerdanyola, Barcelona, España
e-mail: juanantonio.barcelo@uab.es

identifying activity areas, material accumulations or buildings at the site, interpreting burials or settlement patterns can be considered to be within the purview of the analysis of visual marks. There are also non-visual features characterizing ancient objects and materials (i.e., compositional data based on mass spectrometry, chronological data based on radioactive decay measures, etc.), but visual perception makes us aware of many fundamental properties of material evidences of human action in the past.

Unfortunately, there is no universal method of searching for informative visual marks. They can be extracted from any archaeological record almost *ad infinitum*, but one usually fails to formalize the significant criterion for what is intrinsically "visual". An additional difficulty is that different visual features will almost definitely be of importance for different explanations (Shelley 1996). To cope with this problem, archaeologists have traditionally assumed that there is a roughly fixed set or vocabulary of "supposed" descriptive visual regularities shared by a single population of objects, which are also distinctive enough. Archaeologists believe that what they see is a "seed", a "bone", a "bowl", "a knife", the "wall of a house", a "prince burial", etc., and they can distinguish between different kinds of "bowls", different kinds of "prince burials", and so on. This way of identification-based explanation seems then a tricky way of solving any archaeological research problem. It pretends to explain what has been "seen", not in terms of their visual characteristics, but in terms of subjective recognition. Nevertheless, what an archaeologist "sees" at the archaeological site are not stones, walls, pit holes, mounds, buildings, pottery sherds, plants, animal carcasses, or anything like but a hierarchized organization of visual marks and higher level cues to explanatory categories.

The approach we adopt here is to follow current computational theories of visual perception to ameliorate to way archaeology can deal with the analysis and explanation of visual marks. Computer vision has been defined as a process of recognizing elements of interest in an image, and it can be described as the automatic logical deduction of structures or properties of the three-dimensional objects from either a single image or multiple images and the recognition of objects with the help of these properties (Kulkarni 2001). Any reasonable sophisticated visual system must involve a set of processes that extract a variety of types of information from the visual input. This information is captured in a variety of internal intermediate-level representations (neural networks, for instance) which form the basis for higher-level recognition processes.

Following modern studies of computer object-recognition (Grimson 1991; Palmer 1999; Bernardini and Rushmeier 2002; Forsyth and Ponce 2003; Carbonetto et al. 2004; Ponce et al. 2007), we should consider specialized archaeological perception essentially as building larger and larger explanatory structures from elementary visual features. Archaeological explanation is then a gradual process that proceeds from the general to the specific and that overlaps with, guides, and constrains the derivation of a causal explanation from an image or visual representation of some archaeological evidence (Fig. 5.1).

Consequently, it is common to categorize visual process into low, intermediate, and high levels (Marr 1982; Palmer 1999). Low-level information is typically about

Fig. 5.1 A schema showing the process of visual interpretation

Raw visual data

⇩

Image segmentation

⇩

Low-level interpretation of picture elements

⇩

Object recognition

⇩

High-level image interpretation
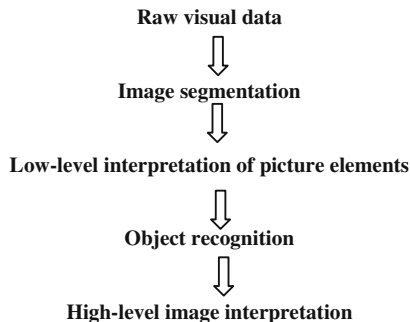
the spatial relationships among primitive, two-dimensional visual features such as observed shape, texture, and composition variability patterns. Intermediate information describes the properties arising from forms of organization of the low-level primitives, and may include descriptions of the three-dimensional spatial relationship (location) among visual properties. The overall explanatory process is thus broken down into the extraction of a number of different observable physical properties (low-level analysis), followed by a final decision based on these properties (high-level analysis), what implies breaking down the perception of meaningful visual marks into different explanatory stages.

The job of the archaeologist is not to provide with a representation of the past in abstract but to look for the information he or she needs to interact with its possible explanation. Visual features should be treated as evidence and their estimation accuracy should be directly correlated to their power to resolve alternative hypotheses. By this account, hierarchies of feature detectors should be combined together in ways given by coactivity of the underlying detectors and necessary knowledge structures necessary to integrate them. Complex association structures are formed when simple feature detectors and prior knowledge structures become associated through repeated sequential fixations of the corresponding features (Barceló 2008).

What is Shape?

The attempts at defining the term *shape* usually found in the related literature are often based on the concept of all the properties of a configuration of points which are not altered for effects of size, position and orientation, or by translation, rotation and scaling (Kendall 1977; Kendall et al. 1999; Bookstein 1991; Small 1996; Palmer 1999; Dryden and Mardia 1998). While such definitions manage to capture an important property of some visual features as perceived by humans, namely what relates the different appearances of the same object seen from different perspectives, they do not clearly specify what a *shape* is. An alternative and less conventional definition of shape has been advanced by Costa and Cesar (2001, p. 266): a shape can be understood as any "single", "distinct", "whole" or "united" visual entity.

When we see something, we are not seeing an object, but our senses capture sensorial information (luminance contrasts), which should then be transformed into an intermediate-level representation of what gives the perceived entity its individuality. Formally speaking, a *surface* is a boundary of separation between two *phases*. In its turn, a *phase* is a homogenous mass of substance, solid, liquid or gas, possessing a well-defined boundary. When we have two phases in mutual contact, we have an *interface*. What gives individuality to any solid entity, kept in atmosphere, is in fact its air-solid interface, or in the case of solid entities in contact, a solid-solid interface, which are often simply referred to as a solid surface. The surface of solids plays a significant role to discover the way they have been produced and the way they have been used.

Surfaces have two main properties: *shape* and *texture*. *Shape* can be best characterized as the perceived interfacial boundaries or discontinuities themselves. In fact, we usually take the geometry of the identified contour or silhouette as a surrogate of the object's shape. It will imply essentially the operation of detecting significant local changes among luminance values in a visually perceived scene and its translation into a geometric language, joining points with lines, fitting surfaces to lines, or "solidifying" connected planes (Barceló 2000). *Texture* is the definition of surface attributes having either visual or actual variety, and defining the appearance of the surface. Any surface has variations in its local properties like albedo and color variations, uniformity, density, coarseness, roughness, regularity, linearity, directionality, direction, frequency, phase, hardness, brightness, bumpiness, specularity, reflectivity and transparency (Tuceryan and Jain 1993; Fleming 1999).

In both cases, geometry is used as a visual language to represent a theoretical model of the pattern of contrast and luminance, which is the strict equivalent of perceptual models of sensory input in the human brain (Barceló 2001). In fact, shapes are concepts corresponding to geometrical abstractions that may never be perfectly represented in the real world. The constructed geometry of an archaeological artifact refers to the idealized form represented by those portions of the artifact that were deliberately modified as part of the production of the artifact from raw material. By idealized geometry (or geometric abstraction) is meant a smoothed form of the interfacial boundary for which variation from the smoothed form appears to simply reflect variation due to the production process (Read 2007).

This implies to consider perceived variation in the interfacial boundary of an artifact to arise from successive modification by the artisan through a sequential, conceptual process going from an initial abstract ideal form to the final geometry of the set of surfaces defining the finished artifact (Van der Leeuw 2000). The particular morphology of the boundary may be determined from physical constraints acting on the process underlying its formation process (artisan work, user action). An example would be the distribution of forces acting on the formation of the boundary of the artifact, as occurs with the hands of the potter making pottery with a pottery wheel. On the other hand, the design of a thrusting spear point is likely to be squat and short with a wide tip angle, combining relatively long cutting edges with a short blade and a relatively wide base suitable for hafting with a strong, robust shaft. A throwing spear point, in contrast, needs to optimize the requirements for aerodynamics, killing power and accuracy. A slim, elongated point combines mass

with a relatively acute tip angle and a small presentation area and base. A smaller base means that a smaller shaft leads to a lower overall weapon mass. According to Newtonian mechanics, a lighter missile can be launched at a higher velocity with a flatter trajectory resulting in a faster, more powerful projectile weapon (Christenson 1986; Flenniken and Raymond 1986; Crompton 2007). Nevertheless, in most cases the underlying physics may be too complex to model if there is no single pattern that constrains the interfacial boundaries.

By virtue of the properties of the raw material and the features of human labor or action, many objects from the past have a constructed shape. This can be the case of tools, pottery containers or most built structures (pit holes, graves, walls, buildings). However, the precise relationship between shape and formation processes is not always direct and easy to explain. In the case of prehistoric stone arrow point made of retouched flint, for instance, its shape of the tool is simply the mechanical consequence of the flake removal, in such a way that the edge of such tools does not represent *necessary* a cognized shape on the part of the artisans (Bisson 2000; Collins 2008).

In some other cases, the actual geometry of perceived interfacial boundaries can be the result of taphonomic processes. The actual shape of a wall, as it is perceived in the moment of the archaeological excavation, is the result of the destruction of the original wall, in such a way that the original ordering of building blocks may be lost. The same is true for a broken pottery vessel, transformed into an amount of fragments whose individual shape is not any more the result of human labor in the past. Mounds resulting from the accumulation of stones, debris or animal bones also can be defined in terms of edges and boundaries explaining the formation (or deformation) processes involved (Mameli et al. 2002). As a result, the precise shape of any archaeological deposition should be analyzed to understand the formation process of the archaeological site (Barceló et al. 2003, 2009). At higher perceptual scales, in the case of soil and landscape features, as territories, valleys, drainage basins etc., the geometry of their interfacial boundaries may also be the result of natural processes or social events having contributed to its actual appearance.

In general, and following Leyton (1992, p. 73) if *boundaries* (or edges) are understood as perceived discontinuities or asymmetries generated through time, we should be able to recover the history of the perceived (and "differentiated") archaeological entity from the perception of change. In other words, archaeologists use shape information, that is to say data on geometric discontinuities, as memories of process-history.

In many other categories of archaeological evidence, the processes of contour formation and transformation result in the essential properties of size, mass and shape changing and reducing with successive re-formation events. However, when dealing with characteristically uneven and asymmetric objects like prehistoric stone tools, built structures and the like, the very concept of shape regularity acquires another dimension, given the particular way an irregular contour may be the result of a sequence of events, modifying each one a previous shape. Central to this concept is the manner in which irregularly shaped archaeological evidences were designed, reduced, resharpened, recycled, and discarded within its use life. In the case of prehistoric stone tools, many of them suffered multiple steps or stages of production.

As the flake tool edge becomes dull, it can be resharpened by the removal of minor stone chips from the dulled area. The more a flaked tool is used and subsequently retouched the greater the amount of visible resharpening either in the form of total length of edge resharpening or total surface area with flakes removed. The same is true for a wall made of building blocks, an irregular mound made of accumulated debris, or an excavated pit hole.

Low Level Visual Analysis

The archaeological record is not made of shapes. It is a series of perceptual information waiting for an observer. The observer will impose order by recognizing interfacial boundaries between different components with different visual marks and by creating a functional model of them. This key assumption has been traditionally neglected in archaeology, preferring a subjective approach where shapes exist as primitives bits of information and may be defined by universal picture stereotypes, e.g., "round", "ovoid", or even worst by user-defined stereotypes, like "hat-shaped," "cigar-shaped," "kidney-shaped". Such assumption does not take into account that objects such as hats and cigars come in a wide variety of morphological configurations, making the visualized reference standard and qualitative terms subject to variations of individual perception. It has been considered an alternative the identification of underlying geometries in qualitative terms, such as irregular, indented, sinuous, etc. to describe it. Even the simple call to standard shapes of Euclidean geometry (rectangle, parallelepiped, circle, sphere, cylinder, cone, etc.) is a misleading answer. Euclidean geometry with its well-defined and mathematically tractable curves and lines is usually only found as an approximation over a range of dimensions where human manufacture labor has imposed it, or in limited situations where a single energy or force term dominates (e.g., surface tension). The surface of artisan-made materials, objects, tools, accumulations or built structures may seem Euclidean only at some particular scale. Magnify the field view and they become rough or irregular.

The result is lack of replication among experts and arguments over the reality of perceived visual information (Dibble 1997; Dibble and Chase 1981; Whallon 1982; Djindjian 1993; Orton et al. 1993; Andrefsky 2005; Read 2007).

Since the observer arbitrarily constructs such configurations, archaeological objects cannot fulfill the parameters of a prototype as long as material objects are governed by the physical variation intrinsic to the labor process that generated the object in the past, and the remaining variation generated by the post-depositional processes that altered its visual characteristics since then.

Visual Data Acquisition and Encoding

In order to be able to increase the objectivity of visual information, archaeologists need a kind of instrumental "observer" equipped with *range* and *intensity* sensors

(Barceló 2005). The former acquire range images, in which each pixel encodes the distance between the sensor and a point in the scene. The latter are the familiar digital cameras acquiring grey-level images. That is to say, such an instrumental observer or automated archaeologist may use a CCD camera to observe a pattern of structured light projected on the scene (e.g., a beam of laser light, or a grid of lines). If the sensor has been calibrated, depth can be inferred by triangulation. In so doing range sensors can measure depth at *single points*, on *lines* (acquiring range profiles), or in 2D *fields of view* (acquiring range images). Most measurements can be directly calculated from a pixel-based representation by simple counting procedures multiplied by calibrated pixel size with adjustments for specific measurements. This measurement is methodologically and fundamentally different from standard measurements based on conventional tools like calipers and tapes.

Digital images contain all the useful information to derive geometry and texture for a 3D modeling application. However, the reconstruction of detailed, accurate and photo-realistic 3D models from images is a difficult task, in particular for large and complex archaeological evidences (Manferdini et al. 2008). Image-based methods require a mathematical formulation (perspective or projective geometry) to transform two-dimensional image measurements into 3D coordinates.

There are three categories of optical 3D data acquisition: (1) image-based methods, e.g. photogrammetry, (2) range-based methods, e.g. laser scanning, and (3) combinations of both. The choice of the most appropriate technology for a given task depends on the object or area under investigation, the experience of the user, the available budget and time, and further parameters. Techniques applied to the restitution of small archaeological objects are based on the exhausting calculation of 3D point clouds, which represent the outer surfaces of the objects. The most popular are:

1. Laser scanning. The measuring of the 3D points coordinates is implemented through a laser beam that is transmitted towards the object and reflected back to the source. The time that is needed for the beam to travel from the laser beam source to the object and back, multiplied by the speed of laser light, yields the distance of the points from the source; hence, their location on an arbitrarily defined 3D coordinates system.
2. Optical scanning. Special structured light devices and laser diodes producing straight (horizontal or vertical) line tracks are used for the exact definition of 3D points on the object. Sophisticated photogrammetric procedures may lead to the calculation of a dense point cloud that describes the outer surfaces of the objects.

These kind of instrumental observers generate as an output detailed point clouds of three-dimensional Cartesian coordinates in a common coordinate system. The digitized data generated by the scanner is composed of thousands of *x, y, z* coordinates that describe a point cloud that represents the surface of the object scanned. The laser scanner measures, in principle, the distance to a target point and the respective vertical and horizontal angles. Besides target distance, the relative intensity of the returned echo signal as well as the true color of the target point should

be recorded in order to obtain an estimation of visual marks variability to be translated into a shape model. A laser digitizer, for instance, captures surface data points that may be less than 300 microns (0.3 mm) apart, producing high-density geometric meshes with an average resolution of over 1,000 points per cm^2. The accuracy of the measurements derived from the acquired point clouds coordinates usually exceed those possible using traditional 2D tools such as calipers and rulers (Doi and Sato 2005; Tsioukas et al. 2004; Trinkl 2005; Kampel and Sablatnig 2006; Petersen et al. 2006; Lambers et al. 2007; Lambers and Remondino 2007; Farjas and García Lázaro 2008; Karasik and Smilansky 2008; Manferdini et al. 2008; Avern 2010, in press).

An alternative approach is computer tomography (Casali 2006; Dimitrov et al. 2006; Kardjilov et al. 2006). The word "tomography" derives from the Greek *tomos* (slice) and *graphein* (to write). Here, the "instrumental observer" scans thin (virtual) slices of the object with a narrow *x*-ray beam, which rotates around the object, producing an image of each slice as a cross section of the object and showing each of the possible internal components in a 10 mm slice. Digital geometry processing is used to generate a three-dimensional image of the inside of an object from a large series of two-dimensional X-ray images taken around a single axis of rotation. The image is made up of a matrix of thousands of tiny squares or pixels (65,000 pixels in a conventional image). Each pixel has an associated measure of how much of the initial *x*-ray beam is absorbed by the different components of the object's fabric at each point in its solid body (the computed tomography number, measured in Hounsfield units). This varies according to the density of the component. The denser the component is the higher the computed tomography number, ranging from 1,000 HU (air) to 1,000 HU (bone). The resulting visual model would be displayed with a different shade of grey for every different computed tomography number. By convention, high computed tomography numbers are displayed as white and low as black.

3D microscopy is another possibility for encoding the microtopography of a surface, and hence some details of its shape geometry. This technology creates a series of individual image planes (up to 200) and overlaps focus levels to construct a three-dimensional composite image (Bello and Soligo 2008).

Using any of those technologies of data capture and encoding, the resulting data is only a spatial array of visual bindings which can be subdivided into sets of marks (points, lines, areas, volumes) that express the position and the geometry of perceived boundaries (shape), and retinal properties (texture).

Edge Detection

Shapes are not something to be captured using digital cameras, laser scans, or computer tomography equipment because they are not a part of reality. We have to "discover" in some way an explanatory representation of luminance regularities that have been acquired and encoded by the "instrumental observer". A geometrical model showing how interfacial boundaries between luminance areas are related

should provide the keys for detecting individual bits of reality in what is apparently a continuous array of visual marks.

The method for "finding" the interfacial boundaries that allow the identification of individualized archaeological observables can be approached by calculating the *luminance gradient* in the data array – that is, the direction of maximum rate of change of luminance values, and a scalar measurement of this rate. It should coincide with the outer frame of the observed object, usually called *edge, contour* or *silhouette*. This is a line marking a constant level of luminance. As an interfacial boundary, this line never ends, although it may branch or loop back upon itself. It is not an intrinsic property of observed objects, but it arises in images in different contexts: discontinuity in the surface depth, discontinuity in surface orientation, markings on the surfaces, etc. In other words, it is the boundary that delimits distinct spatial areas which appear when visual appearances are "significantly different" from one area to the next.

A contour or edge is simply a linear separation between regions with different texture (visual or retinal properties) within an image. Shape discovery is then the operation of detecting significant local changes among luminance values in a visual scene. The method for "finding" edges in the images that represent archaeological evidences can be approached by calculating the *texture gradient* (usually a "luminance gradient") in the data array – that is, the direction of maximum rate of change of luminance values, and a scalar measurement of this rate. Marr and Hildreth (1980) initially defined the procedure, finding the position of maximum variation in the map of luminance (grey or RGB-color levels). First-order differential operators compute the variation levels of such intensity function, and the algorithm finds the edge by detecting the highest value in the first derivative of the intensity function. A more economical algorithm for finding edges would be to detect *zero-crossings of the second derivative* of the intensity function. The second derivative of a function is just the slope of its previously calculated first derivative. The second derivative thus computes "the slope of the slope" of the original luminance function. Notice that in this second derivative function, the position of the interfacial boundary corresponds to the zero value in between a highly positive and a highly negative value (Sonka et al. 1984).

Many different variants and ameliorations of this primitive procedure have been published. The *Canny edge detector* first smoothes the image to eliminate and noise; it then finds the image gradient to highlight regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum (non-maximum suppression). Finally, the gradient array is further reduced by hysteresis to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero (made a non-edge). If the magnitude is above the high threshold, the presence of an edge at this point is affirmed. And if the magnitude is between the 2 thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient above this threshold (Russ 2006).

This is not the proper place to discuss all approaches to edge detection. There is huge literature, indeed an industry, concerned with such algorithms (Martin et al.

2004; Heideman 2005; Russ 1990, 2006; Gonzalez and Woods 2007; O'Gorman et al. 2008). Nevertheless, conventional edge extraction techniques, being sensitive to (image) noise and intensity variations, often *do not* give us the true boundaries of objects in images. On the other hand, their outputs usually contain spurious or weak edges. It is now generally acknowledged that, without a higher-level information of the object itself (such as the geometry of the surface), such techniques produce erroneous results. Consequently, it seems a good idea to build more optimal edge detector by training a neural network with a certain predefined network structure with examples of edge and non-edge patterns. Good examples of this procedure have been published by Wang et al. (2000) and by Martin et al. (2004). The idea is to regard edge detection as a form of statistical pattern classification, using features extracted from an image patch to estimate the posterior probability of a boundary passing through the center point. Only two classes of pixels, edge or non-edge, need to be discriminated. The neural edge detector can directly estimate the probability by training.

Preliminary Approaches to Shape Encoding

Traditionally, archaeologists have referred to diameters and heights when they spoke about shape, forgetting important parameters like surface area or volume. The conventional method for capturing artifact morphology has been to take linear measurements with calipers at fixed loci along an arbitrary line of maximum bilateral symmetry, generally defined as Length (DeBoer 1980; Pobelome et al. 1997; Lycett et al. 2006; Steine 2005; Rushmeier et al. 2007; Mara and Sablatnig 2005a). Such linear measurements, however, are absolute quantities reflecting only size. No geometric information is provided on the relative position of the various breadth and thickness measurements. Consequently, the variables sampled constitute an abstract collection of relative size measurements, approximating the artifact's morphology (see discussion in Crompton 1995, 2007; Meltzer and Cooper 2006).

Size is a magnitude causing all the metric variables to increase in dimension as it increases. On the opposite, shape should be dimensionless, that is, size independent. Obviously, there is no assurance that two archaeological artifacts with identical size values at different parts of their extension will have similar shapes. The shape of every square, for example, is the same whether it is a large square or a small square. An important corollary of this is that attempts to examine shape differences should attempt to account for the effects of isometric size prior to the analysis of shape. To solve this problem, it has been suggested to modify raw size measures that represent a length or width as a proportion of the length of the artifact (Wynn and Tierson 1990; Lycett et al. 2006). An alternative approach would be to average distance or size-based measurements in terms of a global parameter. Feret's Diameter can also be used for this averaging. Generally, it is the greatest distance possible between any two points along the contour. When such a global measure of size is difficult to

calculate, we can estimate a surrogate using a measure of *thread length*. This gives an estimate as to the true length of a threadlike object. It assumes that the object to be measured is threadlike in form.

$$\frac{p + \sqrt{p^2 - 16\,\text{Area}}}{4} \tag{5.1}$$

In Eq. (1), p is the perimeter of the contour, and *Area* is a measure of the surface of the object. Note that this is an estimate only. Use this parameter when the objects are known to be threadlike and bend so that the Diameter parameter is a poor estimate of the true length. Another neglected size parameter is perimeter length. Its measuring, however, often depends to the orientation of the object, and of the kinematics of use of it.

However, dividing each length or width measure by a single measurement designated to represent "size" (such as maximum length, Feret's diameter or height) not always removes correlations with size. Given the problems associated with these methods, geometric mean size-adjustment has been suggested. It implies the size-adjustment of the data on a specimen-by-specimen basis, dividing each variable in turn by the geometric mean of all variables for that individual specimen. This method isometrically corrects for size enabling direct comparison of allometric shape variation (Lycett et al. 2006).

Much more efficient for shape encoding are descriptions based on relational indexes. Russ (2006) gives a preliminary list of the most common and general, apparently adapted to describe any kind of shape:

(1) *Elongation.* Perhaps the simplest shape factor to understand is an Aspect Ratio, i.e., length divided by breadth, which measures an aspect of elongation of an object.

$$\frac{\text{length}}{\text{width}} \quad \text{or} \quad \frac{\text{MaximumDiameter}}{\text{MinimumDiameter}} \tag{5.2}$$

(2) *Roundness.* It measures the degree of departure from a circle of an object's two-dimensional binary configuration. This is based not on a visual image or an estimate of shape; rather, it is based on the mathematical fact that, in a circular object with a fixed area, an increase in the length of the object causes the shape to depart from a circle.

$$\frac{4\,\text{Area}}{\pi p^2} \tag{5.3}$$

In the equation, p is the perimeter of the contour, and *Area* is a measure of the surface of the object. The roundness calculation is constructed so that the value of a circle equals 1.0, while departures from a circle result in values less than 1.0 in direct proportion to the degree of deformation. For instant, a roundness value of 0.492 corresponds approximately to an isosceles triangle.

(3) *Shape Factor* (or Formfactor). It is similar to *Roundness*, but emphasizes the
    configuration of the perimeter rather than the length relative to object area. It
    is based on the mathematical fact that a circle (Shape factor value also equal to
    1.0), compared to all other two-dimensional shapes (regular or irregular), has
    the smallest perimeter relative to its area. Since every object has a perimeter
    length and an area, this mathematical relationship can be used to quantify the
    degree to which an object's perimeter departs from that of a smooth circle,
    resulting in a value less than 1.0. Squares are around 0.78. A thin thread-like
    object would have the lowest shape factor approaching 0.

$$\frac{4\pi \ \text{Area}}{p^2} \tag{5.4}$$

In the equation, p is the perimeter of the contour, and *Area* is a mea-
sure of the surface of the object. Notice that formfactor varies with surface
irregularities, but not with overall elongation.

(4) *Quadrature*: The degree of quadrature of a solid, where 1 is a square and 0.800
    an isosceles triangle. This shape is expressed by:

$$\frac{p}{4\sqrt{\text{Area}}} \tag{5.5}$$

In the equation, p is the perimeter of the contour, and *Area* is a measure of
the surface of the object.

(5) *Curl*. It measures the degree of departure of an object from a straight line,
    which usually is applied to irregular lines or long, narrow (squiggly) objects.

$$\frac{\text{length}}{\text{skeletonlength}} \tag{5.6}$$

In the equation, maximum length (or Feret's diameter) should be divided by
the object's symmetry axis length or its center line distance (skeleton length).

(6) *Solidity*. This measure is based on the ratio of the area of the true object to the
    area of a snug polygonal box fitted around the object. The degree of difference
    between the object and its fitted box is a quantitative measure of the degree of
    irregularity of the object; irregularity itself becomes a quantifiable aspect of
    morphology.

$$\frac{\text{Area}}{\text{ConvexArea}} \tag{5.7}$$

(7) *Convexity*. This measure is based on the ratios of the perimeter of the true
    object to the perimeter of a snug polygonal box fitted around the object. The
    degree of difference between the object and its fitted box is a quantitative mea-
    sure of the degree of irregularity of the object; irregularity itself becomes a
    quantifiable aspect of morphology.

$$\frac{\text{ConvexPerimeter}}{\text{perimeter}} \tag{5.8}$$

(8) *Compactness.* It is defined as the ratio between the length of the object's contour (the perimeter) and the perimeter of a circle with the same area. It is always greater than 1 and approaches unity when the basin approaches a circular shape

$$\frac{\sqrt{(4/\pi)}\ \text{Area}}{\text{MaximumDiameter}} \tag{5.9}$$

This way of measuring shape parameters is very popular in archaeozoology, paleontology and physical anthropology (Mafart and Delinguette 2002; Rovner 2006; Rovner and Gyulai 2007), but not so common in mainstream archaeology, where qualitative and subjective descriptions are the rule. In the case of lithic tools, Rovner (1993), Russ and Rovner (1989), Rovner (2006) has used such methods to analyze the shape of tools. J.A. Barceló and J. Pijoan (Barceló et al. 2001; Adán et al. 2003; Barceló and Pijoan 2004; Pijoan 2007) have used geometric relational indexes to describe variation in use-wear textures in terms of the shape variability of determinable areas with homogenous micro texture. Beyond the study of objects' shape, Bardossy and Schmidt (2002) have used indexes of compactness, circularity and elongation, for the study of macro-scale entities, like landscape features (drainage basins morphology).

There are many other relational indexes specific of particular categories of archaeological evidences. In the case of pottery, the work by Ericson, Read and Burke (1972) pioneered such approaches, introducing measures based on the calculation of the centre of gravity, to study the relationship between shape and equilibrium, associated with the assumed function of a pottery vase as container. To locate the center of gravity, Bishop et al. (2005) suggest looking for the intersection of the length of the major and minor axis of the ellipse with the same normalized second central moments as that part of the object – i.e., the vase's mouth, its body or its base. As a result, we can measure:

(9) The *Equivalent Diameter* is the diameter of a circle with the same area as this part of the vase, computed as:

$$\sqrt{(4^{*}\ \text{Area}/\pi)} \tag{5.10}$$

(10) The *eccentricity* is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. *Orientation* is the angle (in degrees) between the *x*-axis and the major axis of the ellipse that has the same second-moments as the region. *Solidity* is the proportion of the pixels in the convex hull that are in the region computed as:

$$\text{object's area/convex area} \tag{5.11}$$

(11) *Extent* is the proportion of the pixels in the bounding box that are also in the region, computed as:

$$\text{object's area/area of bounding box} \qquad (5.12)$$

M. Smith (1983) suggested measures like the relative restrictedness or the absolute orifice size. Working on Smith's approaches, K. Juhl (1995) has provided a detailed list with more than 40 relational indexes based on the size differences between different parts of the same artifact. Among them:

(12) Restriction ratio: the surface are of a vessel mouth at its most restricted point divided by the area of a circle with the same maximum radius as the vessel

(13) Relative access factor: the surface area of the vessel's mouth at it most restricted point divided by the volume below this point.

(14) Relative restriction factor: the circumference of the rim divided by the total surface area of the vessel

(15) Leverage factor: vertical distance between the base and the centre of gravity divided by the bottom radius.

(16) Relative position of the centre of gravity: vertical distance between the bottom and the centre of gravity divided by the total vessel height.

Porter et al. (2005) add the following shape parameters for pottery containers:

(17) Relative width at base. This is the maximum width at which the object touches the ground relative to the object's maximum width.

(18) Relative headroom. This is the maximum clearance of a concave base relative to the object's height.

(19) Clay efficiency. This is the ratio of a vessel's capacity relative to the volume of the raw material (clay) used for manufacture.

(20) Relative centre of gravity. This is based on the assumption of a homogeneous density of the clay and set in relation to the object's height.

(21) Relative access width.. This is the width of the inner point with the biggest (inner) height-to-width ratio relative to the maximum width.

(22) Angle of access. This is the maximum angle under which one can directly reach the middle of the vessel's bottom, calculated from the access width.

(23) Mean relative wall thickness. Computed along the dominant skeleton arm (by doubling the minimum distance to the next contour point), and relative to the diagonal of the smallest enclosing rectangle (to incorporate both very wide as well as very tall vessels). Feet and other decorations and ornaments lead to a slightly overestimated value.

(24) Skeleton-complexity. The number of additional skeleton arms (belonging to at least 5 profile points).

Very similar are relational indexes for lithic tools, bronze artifacts, or even in the case of macro-scale entities, like walls, buildings, settlements or even landscape or territorial features. The differences come from the assumed functional relevance of some of the indexes. For instance, in the analysis of the shape of lithic tools, there are:

(25) Index of robustness, calculated as

$$width^*thickness/length \qquad (5.13)$$

Higher values imply more robust tools.

(26) Index of reactivation, calculated as:

$$Length/thickness \qquad (5.14)$$

(27) An alternative measure for the reactivation index is an estimate of the quantity of modification affecting the original shape of the tool generated by the successive removing of flakes from a blank core. Kuhn's (1990) geometric index is:

$$height\ of\ retouch/blank\ thickness. \qquad (5.15)$$

(28) Clarkson's "index of invasiveness" assumes the tool begins its use life as an unmodified flake, and gradually as the tool undergoes use, it also undergoes resharpening and retouching. The measure is based upon the relative proportion and size off take scars to the unflaked surface of the tool blank. A tool with two completely flaked surfaces would have the highest retouch value and a tool with no flakes removed from its surface would have the lowest retouch value (Clarkson 2002. See also Andrefsky 2006).

Many other relational indexes are possible, notably in the case of arrow or spear points. In these cases, angles serve as a major shape descriptor.

Measuring the degree of symmetry is another shape parameter that has relevance for studying the form-function relationship of an archaeological artifact. A very simply approach would imply to divide width measurements into a left and right measurement, and measuring them along the major axis of the object divided into ten or more segments. The idea is to consider changes in the object's width in equally spaced intervals increments down its length. When displayed as a histogram, changes of width of the point appears as bars above (expansion) and below (contraction) a centerline. Families of similar points should produce repeated and recognizable patterns in the change in width, with respect to length (in the case of lithic tools: Dibble and Chase 1981; Morris and Scarre 1981; Henton and Durand 1991; Lycett et al. 2006, in the case of pottery objects: Wilcock and Shennan 1975; Richards 1987). In such a way, the degree of curvature – a shape parameter- of the outline can be estimated. An alternative is based on a relational index expressing the degree of asymmetry between two bilateral measurements as a ratio of the overall width. Lycett suggests computing

$$S = \sum_{i=1}^{n} \left(\frac{\sqrt{(xi - yi)}^2}{xi + yi} \right) \qquad (5.16)$$

where x_i is the width value left of the length line taken at a particular percentage point, y_i is the width value right of the length line taken at the corresponding percentage point and n is the number of percentage points at which x_i and y_i are taken. Hence, a value of zero would correspond to perfect bilateral symmetry (Lycett 2008; Lycett et al. 2006. See also Hardaker and Dunn 2005).

Edge Curvature as Shape Encoding

To completely characterize a shape means to be able to re-create the shape using only the measurements made over the shape. D. W. Read formalizes this requirement in the following definition: "An ordered n-tuple of measurements completely characterizes a shape without redundancy if (a) there is a set of drawing rules that permits reconstruction of the shape outline using only this ordered n-tuple of measures, and (b) there is no ordered k-tuple of measures, $k<n$, such that the shape outline can be reconstructed from the ordered k-tuple" (Read 2007: p. 157). We can also refer to a set of measures that completely characterizes a shape as satisfying the archival property: the shape can be reconstructed from the measures that have been taken. The archival property is a weaker requirement that non-redundancy in that a set of measurements satisfying the archival property may possibly be a redundant set of measures. Additionally, it should be a measure (or series of measures) not changing under similarity transformations: translations, rotations, and changes of geometric scale (enlargements or reductions).

The general characteristics of contour or silhouette of a solid is a good candidate for a complete shape descriptor following Read's conditions. It is usually defined as the longest elongation around – or cross-section through – the outer limit of the object defined by its rotational axis (axis of symmetry) (Mara and Sablatnig 2008). It provides a relatively compact way of representing the shape of an object, with the assumption that the region between the edges defining the contour is relatively homogenous.

There are many ways of considering the overall geometry of contour invariant to size, scale and transformation. A relatively simply approach would be using a form of run-length encoding (also called chord encoding). This treats the image as a series of scan lines. For each sequential line across each region of feature, it stores the line number, start position, and length of the line. A simple polygonal approximation to the boundary can be produced from the run-length table by using the endpoints of the series of chords. A special form of this polygon can be formed from all of the boundary points, consisting of a series of short vectors from one boundary point to the next. On a square pixel array (a bit-map image), there are only eight possible directions. Assigning a digit from zero to seven to each direction and writing all of the numbers for the closed boundary, we will produce a chain code representing shape (Russ 2002; pp. 373–375). This approach was used in some of the first essays of shape analysis in archaeology (Kampffmeyer et al. 1988; see also Hagstrum and Hildebrand 1990).

The way G. Laplace described the operative edge of prehistoric lithic tools can also be classified within this group of applications (Laplace 1972). He advocated for an objective description of induced modifications along the cutting edge ("retouches") based on qualitative identifications and quantitative properties like mode (angle), amplitude, orientation, linearity. A combination of qualified retouches along the edge allowed a semi quantitative description of the edge's shape.

A much more efficient alternative implies converting the contour directly into a mathematical representation of the digitized boundary. In some cases, the representation may be achieved through characterizing the boundary with a mathematical expression such as a polynomial equation. A polygon can fit a contour line through the points interpolated between the raw image pixel centers for all such pairs of pixels that bracket the contour value. This mathematical expression also serves as the rule for drawing the curved segment from the parameter values. The main difficulty in implementing this method for representing a boundary lies in identifying the appropriate mathematical expression. Smoothing will eliminate variation that represents the idiosyncrasies of artifact production and individual variability. For a smooth curve, parameter estimation using statistical curve fitting methods will be quite accurate and constructing confidence intervals for parameter estimations will not be needed. The size of the confidence interval is likely to be on par with or smaller, than measurement error introduced though digitization (Arlinghaus 1994; Dierckx 1995; Hermon et al. 2001; Schurman et al. 2002; Sarfraz 2007).

Polynomial expressions (or their equivalents: Bezier curves/B-splines) will provide a mathematical expression for virtually any curve that can be interpreted as representing the geometry of an object's contour. Polynomial interpolation has a long history in archaeology, especially in the case of pottery studies. Hall and Laflin (1984) used B-spline methods to convert digitized pottery containers silhouettes into one or more mathematical curves (see also Smith 1985; Kampel and Sablatnig 2003b; Mom 2005, 2006; Nautiyal et al. 2006). Since the interpolated profile curve is a planar curve, we can assign a sign to the curvature: i.e. positive or negative (Simon et al. 2002). If the curvature at each point on the object's contour is defined by constructing the circle, which osculates the curve at the point of interest, then the curvature will be the inverse of the circle's radius, and it will be positive if the curve is convex at that point and negative when it be concave. In this way, we can obtain additional shape qualifiers. For instance, a contour can be concave, convex, or planar. A concave edge is characterized by inflection points where the adjacent sections of the curve form an angle of less than 180 degrees. A convex edge is characterized by inflection points where the adjacent sections of the curve form an angle of more than 180 degrees. A planar edge is characterized by inflection points where the adjacent sections of the curve are coplanar (Jang et al. 2006).

When constructing interpolating polynomials to represent boundary curves, there is a tradeoff between having a better fit and having a smooth, well-behaved fitting function. The more data points that are calculated in the interpolation, the higher the degree of the resulting polynomial, resulting in greater oscillations between data points. Therefore, a high degree interpolation may be a poor predictor of a function between points, even though the accuracy at the data points will be "perfect".

Quantitative measures of the *intensity* of curvature should also be calculated to enhance the global understanding of the geometry allowed by the polynomial expression of the contour. The rational behind this approach comes from a differential geometry theorem, which states that any two curves which have identical curvature and torsion are the same curve regardless of translation and rotation (Lu et al. 2007).

One of the earliest methods is the "tangent profile" or its later development the "sampled tangent-profile" technique (Leese and Main 1983). The curvature *k* of a planar curve, at a point on the curve, is defined as the instantaneous rate of change of the slope of the tangent at that point with respect to arc length (for the general procedure, see Bebis et al. 1998; Heideman 2005). The curve is described by providing the tangent as a function of the arc-length, since the curvature is the first derivative of the tangent angle. A related method was used by Liming et al. (1989), who expressed the shape by providing the distance of the points on the profile from the axis of revolution as a function of the arc-length.

Another relatively simple measure of "surface curvature" can be calculated by taking the standard deviation of the *z*-coordinates along the *x*-axis and the *y*-axis, and dividing this by the length of those. Hence, the coefficient of surface curvature emphasizes relative variation over the length of each axis. In the same way, a "coefficient of edge undulation" can be estimated by computing the standard deviation of the *z*-coordinates at the endpoints of the length, and dividing this by the geometric mean of the lengths of the four axes (Lycett et al. 2006).

Additional measures of edge curvature are possible. Consider the coordinates of each point on a curve *x(s), y(s)* where *s* denotes the arc length along the curve. As the parameter *s* changes, the point moves along the line. *x(s)* can be chosen as the distance from the axis of cylindrical symmetry of the vessel. At each point on the curve, the tangent vector can be drawn, which by convention, points in the direction of increasing arc length along the curve. Denoting by θs the direction of the tangent vector with respect to a fixed axis (for simplicity the axis can be chosen as the x-axis), then the tangent angle $\theta(s)$ will determine the curve. The curvature $\kappa(s)$ will measure the rate of change of the tangent angle:

$$\kappa(s)) = (d\,\theta/ds) \tag{5.17}$$

An alternative definition can be given in terms of the radius $\rho(s)$ of the circle that osculates the curve at the point *s*:

$$\kappa(s) = 1/\rho(s) \tag{5.18}$$

The advantage of this representation is that $\kappa(s)$ is large where the line changes its direction in the most rapid way, which are also the points of greatest interest in many archaeological applications. Additionally, a small indentation that only changes the curve locally will appear as a small perturbation in these representations. The tangent angle depends more strongly on local features (it is defined as a ratio of derivatives of the Cartesian coordinates), and hence, local changes of the

line will show up. The curvature involves a higher derivative, and thus it is very sensitive to local variations. The features of the line that provide information on the gross properties of the curve will be hardly shown. The selection of the representation to be used is dictated by the particular application at hand, and on the features of the curve which are of relevance (Gilboa et al. 2004; Saragusti et al. 2005; Karasik 2008; Nautiyal et al. 2006; Mom 2005, 2006; Maaten et al. 2009).

More details about the geometry of a curvature can be measured on a digitized profile. In the case of microtopographic features of bone surfaces, Bello and Soligo (2008) suggests measuring the following parameters (Table 5.1). Although specifically oriented to micro-shape analysis, they can be easily generalizable to the curvature of any kind of profile.

This way of shape encoding allows us a different way to calculate "shape regularity". In this case, it is meant as a measure of the variations in corners, edges and faces, or, simply a measure of directional changes in the object's contour. Consequently, and following Saragusti et al. (2005) we can quantify the degree of roughness of a given contour based on the degree of concavity of this contour. The intuition at the root of this measure of roughness stems from the following observations: the smoothest closed curves are convex. Any further structure of the curve is associated with the appearance of concave sections: the more there are the more complex and rough the curve is. Thus, roughness can be determined by the frequency and amplitude of the transitions between convex and concave sections along the curve. These transitions occur at inflection points. The concavity can be defined as the sum of all the deflections along concave sections. It should be borne in mind that roughness is a relative term, and it depends upon the scale at which it is defined and measured. A given line may look

Table 5.1 Curvature advanced descriptors (from Bello and Soligo 2008)

- *Slope angles* ($\sigma 1$ and $\sigma 2$): the angles between the slopes S1 (left) and S2 (right) of the cutmark and the unaffected bone surface (R).
- *Opening angle of the cutmark* (δ): the angle between the slopes S1 and S2 ($\delta = 180° - [\sigma 1 + \sigma 2]$)
- *Bisector angle* (BAC): angle of the bisector of the opening angle of the cutmark relative to the unaffected bone surface (expected to reflect the impact angle of the tool relative to the bone surface; $\gamma = \sigma 2 + (180° - [\sigma 1 + \sigma 2])/2$)
- *Shoulder heights* (SH, left and right): the height of the shoulders formed on either side of the cut (SH = $\sin \beta$ L, where L is the distance from the tip of the shoulder to the corresponding intersection between the cutmark profile and regression line R, and where β is the angle between L and R).
- *Floor radius*: the radius of a circle fitted to the floor of the cutmark profile, with the floor defined as lying between the two points where the profiles of the left and right slopes start to converge (i.e., where the cutmark profiles start to diverge from the regression models S1 and S2).
- *Depth of cut* (DC): the perpendicular depth of the cut relative to the unaffected bone surface (DC = $\sin \alpha$ - H, where H is the distance from the lowest point of the cutmark profile (point A) to the intersection between the left slope of the cutmark profile and the regression line R (point B), and where a is the angle between H and R).

relatively smooth at one scale, and rougher as the resolution increases. Therefore, setting the scale at which the roughness is to be measured is a prerequisite in any quantitative assessment of its degree. Archaeological considerations dictate the choice of scale, and may change when different properties are to be addressed. It is convenient to set the scale by assessing the size of an arc along the curve, within which variations of the curvature are irrelevant and can be smoothed away. The fluctuations of the curve that occur on an interval of smaller length are damped out.

Alternative Descriptors of Global Shape

The method of archaeological curvature estimation and measuring is not without its limitations. One limitation lies in the requirement that (a) the curve be placed in a rectangular coordinate system in such a way that a line vertical to the x-axis will intersect the curve at most once and (b) the curves to be compared to each other have a comparable way for orienting them in the coordinate system. Closed curves, clearly contradict the first requirement. However, if the outline is divided into segments based on the corners of the object outline, each segment can be oriented with its own coordinate system. On the other hand, if the orientation of the object's profile is changed, then the polynomial curve parameters will assume different values even though there is no change in the form of the curve (Mara and Sablatnig 2006). One needs a "natural" orientation of the curves with respect to the coordinate system, and whether there is, a natural orientation depends on how the archaeological artifact was conceptualized and produced in the past, and the level of its preservation in the present. It is perfectly possible that the object has no predefined orientation because the shape responds to local attributes. Re-sharpening a lithic tool by subsequent flake removal (retouching) would be an example (Read 2007). Additional problems have been remarked by van Maaten et al. in the case of uncentered objects with excessive wear, irregular shape and/or edges, deterioration, and so on (Maaten et al. 2006; see also Mom and Paijmans 2008). If an object does not have exact rotational symmetry (e.g. hand shaped vessels, arrow points, flint axes, etc.), one may obtain several different shapes by drawing a single object from different angles, in which case the calculation will produce a measure for the internal symmetry of the object. On those cases, edge-based statistical tend to provide insufficient shape description.

Some of these problems can be solved using alternative procedures for global geometry description. For instance, it has been argued that the perimeter of complex entities may have a fractal nature. A fractal analysis of shape should therefore an alternative that merits to be explored.

The *fractal dimension* is a statistical quantity that gives an indication of how completely a fractal appears to define an interfacial boundary, as one zooms down to finer and finer scales. Consequently, it can be understood as the rate at which the profile or contour of an object increases as the measurement scale is reduced (Russ 2002; Rovner 2006). There are different ways to measure it. Perhaps the most

widely used fractal measurement tool is the so-called Richardson plot. In analysis of an irregular contour curve, the effects of varying resolution can be mimicked by walking a real or virtual map divider along the curve, varying divider step size to obtain different estimates of curve length. Setting a pair of dividers to a known distance, the user starts at some point on the boundary and strides around the perimeter. The number of steps multiplied by the stride length produces a perimeter measurement. As the stride length is reduced, the path follows more of the local irregularities of the boundary and the parameter increases. A logarithmic plot of results shows degrees of curve wandering at specific scales by the plot slope values associated with particular step sizes. A plot or plot segment that is linear shows statistically self-similar character over the relevant range of scales. In many cases, Richardson plots from natural curves show abrupt shifts in slope with changing resolution. It means the presence of contrasts in curve geometry between two ranges of scale, and suggests scale thresholds in formative process. The slope of the regression line on the plot gives the fractal dimension, and it can be used to measure the relationship between the measuring scale and the attribute of the image being measured (Jelinek et al. 1998).

This method of shape description has been used in the analysis of prehistoric lithic tools by Kennedy and Lin (1988) and by Brown (2001). Fracture lines characterizing the contour of a broken fragment can also have a fractal nature (Pande et al. 1987; Brown and Wiltschey 2003a; Brown et al. 2005; Leitao et al. 2005). Some other relevant applications are in the domain of the shape of spatial features, like territorial limits and the like. In fact, the body of literature in modern geography on the fractal characteristics of the global shape of human settlement is significant and growing. The shape of several different kinds of modern and ancient settlements has been shown to be fractal in form. A number of investigators have studied the boundaries of modern cities and prehistoric settlements and concluded that they are fractal curves that can be modeled by a process called diffusion limited aggregation. In any case, not all settlement patterns should be fractal. For example, the orthogonal grid pattern of an archetypal Roman city tends to be Euclidean rather than fractal, although its fractality depends on the details of the grid squares (Puente and Castillo 1996; Willemin 2000; Bardossy and Schmidt 2002; Brown and Witschey 2003b).

Fractal dimension produces a single number that summarizes the regularity of "roughness" of the contour line. However, there can be an unlimited number of visually different boundary shapes with the same fractal dimension or local roughness. The most usual alternative is then *harmonic analysis*, a shape unrolling method that converts the observed boundary to a function of the angles of radii drawn from the object's centroid until the points delimiting the contour. In both cases, the contour is described in terms of its polar coordinates: a point on the outline is located by the angle from a reference line and the distance from the center of the polar coordinate system (the reference point) along a ray at that angle to the outline. By using a fixed set of angles (even unequal angles), the measurements can be reduced to the length of the rays. The combination of increasing accuracy of the representation through increasing the number of rays or diameters (widths) and then treating each ray or width as a variable has the drawback that it introduces redundancy into the

system of measurements. Two methods of harmonic analysis have been most used in archaeology: Fourier analysis and Hough Transform.

The power of the Fourier representation of a curve lies in the ability to represent even irregular curves simply by including a sufficient number of terms with nonzero coefficients in the Fourier series. The problem is that corners are difficult to approximate with the waveform represented by sine and cosine functions; hence a large number of terms are needed to force the curve represented by the Fourier series to represent a corner. Fourier decomposition has been applied to archaeological data by Gero and Mazzullo (1984), Cardillo (2005), Cardillo and Charlin (2007), to investigate the outlines of flake tools; by Karasik et al. (2005) and Goel et al. (2005) in the analysis of asymmetrical deformations of pottery vases; by Forel et al. (2009) to study the morphometry of European Bronze Age (2300–800 BC) bronze tools; and by Peterson (1992) to estimate shape differences among the limits of particular fields during Roman Times.

Alternatively, transformation into Hough spaces have been used to find and understand alignments of points along a contour and fit some kinds of shapes, although it is necessary in most cases to have a pretty good idea of the type of line or other arrangement that is to be fit to the data. Hough transform algorithms use the polar coordinate representation of the contour consisting of radius length and angle. Each point in the real-space image produces a sinusoidal line in Hough space representing all possible lines that can be drawn through it. Each point in Hough space corresponds to a line in real space. The superposition of the sinusoids from several points in real space causes the lines in Hough space to add together where they cross. These crossing points identify the contours that go through the points in the real-space (Russ 2002). Generalized Hough transforms was one of the first algorithms to be explored for the automatic documentation of pottery profiles (Lewis and Goodson 1990; Durham et al. 1990). Modern applications include Kampel and Melero (2003) on pottery profiles, and Keogh et al. (2009) have used the same approach for describing the shape of petroglyphs and pictographs in rock-art.

Salient Points as Shape Encoding

In many applications of archaeological shape analysis, we need to assume that the shape of any object is effectively captured by a finite subset of its contour points, in such a way that we do not need the entire geometry of the contour for understanding it. Selected points along the contour should correspond to salient points. "Salient" means that the point is in some way "special" or "distinct from its neighbors". Attempts to define what a salient point is suffer from the problem that an isolated point cannot be special by itself, but only in comparison to its neighbors along the object's contour. Hence, saliency makes sense only with respect to the surroundings. In the literature there have been various synonyms for these selected contour points: inflection points, vertices, anchor points, control points, profile points, sampling points, key points, facets, nodes, markers, fiducial markers, and so long. The

most usual way to refer to them is the term *landmark* (Adams et al. 2004; Slice 2007).

Dryden and Mardia (1998) define a landmark as a point of correspondence on each object that matches between and within populations. Landmarks with the same name, homologues in the purely semantic sense, are presumed to correspond in some sensible way over the forms of a data set. These authors consider three basic types of landmarks: functional, mathematical and pseudo-landmarks (or semilandmarks).

(1) A *functional landmark* (or "anatomical", as it is used in paleontology and biology) is a point assigned by an expert that corresponds between similar objects in some explanatory meaningful way. Functional landmarks designate parts of an object that corresponds in terms of functional derivation and these parts are considered homologous. From a production viewpoint, for instance, a corner along a digitized contour and formed by the intersection of adjacent sides involves a discontinuity representing the intersection of two different formation processes (Leyton 1992, 2005; Read 2007; Collins 2008). In pottery studies, the transition between the neck and body of a vessel is a functional landmark (Rowe and Razman 2003). Considering the process of making a pottery vessel from the original amount of clay, Kampel and Sablatnig (2007) have suggested, in the case of pottery a series of functional landmarks (Table 5.2).

In the case of lithic tools, S. Crompton (2007) has selected a series of functional landmarks to describe attributes, which affect the point usefulness as throwing or thrusting spear. Based on these functional principles, Crompton has selected 11 landmarks on specific points of the tool: on the tip, on the maximum length and width, on the hafting edge, etc.
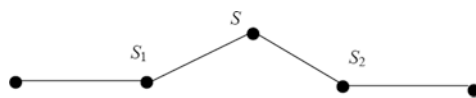
In other examples of lithic tools, it has been argued the impossibility of marking functionally homologous landmarks because of the lack of such knowledge (Cotterell and Kaminga 1992; but consider the opposite arguments in Laplace

Table 5.2 Functional landmarks for pottery studies (from Kampel and Sablatnig (2007))

- SP, starting point: in the case of vessels with a horizontal rim: innermost point, where the profile line touches the orifice plane.
- OP, orifice point: outermost point, where the profile line touches the orifice plane.
- IP, inflexion point: point, where the curvature changes its sign, i.e. where the curve changes from a left turn to a right turn or vice versa.
- MI, local minimum: point of vertical tangency; point where the x-value is smaller than in the surrounding area of the curve.
- MA, local maximum: point of vertical tangency; point where the x-value is bigger than in the surrounding area of the curve; the y-value refers to the height of the object (e.g. MA(y)).
- CP, corner point: point where the curve changes its direction substantially.
- BP, base point: outermost point, where the profile line touches the base plane.
- RP, point of the axis of rotation: point where the profile line touches the axis of rotation.
- EP, end point: point where the profile line touches the axis of rotation; applied to incomplete profiles.

1972). It seems easier in the case of metallic objects, like iron brooches (Small 1996; Le and Small 1999).

(2) *Mathematical landmarks* are points located on an object according to some mathematical or geometrical property of the figure, irrespective of its functional value. It may be a matter of subjective evaluation if such points have or not functional value, given what we have suggested previously about the causal nature of corners and inflection points along a contour (Leyton 1992). Mathematical landmarks can be defined, in fact, as any set of points that are characterizable and searchable upon a surface. They can be based either on geometric properties of profile (maxima of curvatures, umbilic points, crest lines) or, if available on the retinal properties of the underlying images – i.e., color, luminance (Caldoni et al. 2006; Maiza and Gaildrat 2006). A suggested method is based on finding points of inflection on a curve at varying levels of detail (i.e., curvature zero-crossings). In general, points on the outline of a curve where there is an abrupt change in slope are usually considered as control points (Rowe and Razman 2003; Goel et al. 2005). Control points are chosen on the pot's profile by removing the irrelevant shape features and keeping the relevant ones. This is achieved by iteratively comparing the relevant measure of all points on the profile. For each of these iterations, the vertex that has the lowest relevance measure is removed and a new segment is established by connecting the two adjacent points. A point with a higher relevance value signifies that it has a larger contribution to the shape of the curve. Goel et al. (2005) give the following formula to compute the relevance of a point on the curve.

$$K(s) = \frac{|B(s_1, s_2) - 180| \, l(s_1) l(s_2)}{l(s_1) + l(s_2)}$$

S, S1, S2 are points on the profile of a curve
K(S) is the relevance of the point S to the shape of the curve
B is the turn angle of S with points S1 and S2 is the length between S and either of the other points.

Martín et al. (2009) take a somewhat different approach to the selection of salient, by simplifying a scanned point cloud.

(3) Unfortunately, landmarks cannot always be easily defined and located with precision. In any case, such a process could not be made straightforwardly when the complete objects silhouettes do not fully correspond to their original shapes. In some case, the cutting edge may have been drastically reworked due to repetitive sharpening operations accentuating the curvature by plastic deformation. *Pseudo-landmarks* (also called semilandmarks) are constructed points on an object, located either around the outline or in between anatomical or mathematical landmarks. For instance, when we consider equally spaced points along a contour, we can select any points (usually 100 or more), provided

we are sampling the shape with roughly uniform spacing (for the general procedure, see Belongie et al. 2002). M. Cardillo makes the important point that we need a minimum of two mathematical or functional landmarks (for instance both extremes of the symmetry axe) to fix a series of pseudolandmarks for comparison purposes (Cardillo 2006. See also Otárola-Castillo et al. 2008).

An example is the configuration of 51 landmarks employed for the 3D analysis of nuclei surface morphology by Lycett et al. (2006). In the case of pottery, and related with pseudo-landmarks measuring a profile at specific intervals, Maaten et al. (2006) suggest a related technique called shape context. In a shape context representation, a shape is represented by a number of points that is sampled from the boundary of the shape contour. The points are described as shape context descriptors. Shape context descriptors describe the distance and angle of a point to all other points in a discretized log-polar space (see also Maaten et al. 2009).

A good example of the object identification possibilities offered by these procedures for shape encoding has been provided by E.S Lohse and his Archaeological Auto Classification System (Lohse et al. 2004), which uses neural network technology to classify stone arrow points based on the geometrical information of their contours, described in terms of pseudolandmarks.

Geometric Morphometrics analyzes the variability in the relative position of landmark-coordinates said to be "homologous" in terms of the phenomenon being studied. The procedure begins by determining a *configuration* in the set of functional, mathematical or pseudo-landmarks on a particular object. The *configuration matrix X* is the k x m matrix of Cartesian coordinates of the k landmarks in m dimensions. We shall consider a shape space obtained directly from the landmark coordinates, which retains the geometry of a point configuration. The configuration space is the space of all possible landmark coordinates. Consequently, the selection of the appropriate coordination system is of paramount importance. A suitable choice of coordinate system for shape should be invariant under translation, scaling and rotation of the configuration. Among the most common, *bookstein coordinates* are the remaining coordinates of an object after translating, rotating and rescaling the baseline to preserve the original geometry (Bookstein 1991; Dryden and Mardia 1998; Adams et al. 2004; Slice 2007).

Between landmark points, we can define the existence of *paths* connecting any two-neighbor points. In this way, parametric curves are connected sets. Two types of paths are usually considered: polygonal and continuous. A *polygonal path* is defined as a sequence of connected straight-line segments, i.e., straight segments sharing their extremities. In case there is a polygonal path with all its points contained in the set, and linking any two points, then this series of landmark points is *polygon connected*. A polygonal path corresponds to a particular case of a continuous path. A series of landmark points is *path wise connected* in case any two of its points can be joined by a continuous path entirely contained in this particular set of landmark points.

Shape differences between the objects described by landmarks and paths between landmarks are demonstrated by superimposing the landmark configurations according to some criteria or by making them to coincide (Slice 2007). A geometrical morphometric analysis based on Generalized Procrustes Superimposition Analysis (GPA) is applied to remove variation in location, orientation and scale. This procedure specifically addresses shape differences apart from size. After superimposition, GPA creates a "mean" or average shape from which the variability in overall landmark positions can be quantified. This is analogous to scaling and rotating photographic negatives of the samples as represented by the landmarks and superimposing their corresponding landmarks to obtain an overall best fit. After fitting the landmark configurations to the computed mean shape, the shape differences are recorded as residuals from this mean shape.

Shape at the Third Dimension

In all precedent references, shape has been referred as a *bi-dimensional* geometrical parameter, that is, as a curve. However, archaeological evidences, as any material object are three-dimensional entities, and their bi-dimensional contour is but a crude surrogate of their real shape. Interfacial boundaries of archaeological material evidences (ceramic vessels, bones, stone tools, ancient walls, occupation floors, burials or pit holes) have the appearance of *irregular surfaces*, more than *curves*, and we should take into account that due to "the loss of dimension", 2D images of 3D objects suffer from ambiguities.

We may need then more sophisticated mechanisms for analyzing archaeological shapes in all their complexity. As objects become more complex in terms of variety of shape and changes in curvature, they become more difficult to quantify and analyze. We need then more sophisticated mathematical techniques to represent complex surface geometries of intrinsically three-dimensional objects (Hermon 2008). Some of the 2D shape descriptors examined in previous sections can be adapted to be used in the 3D case. Examples of global descriptors would be the statistical moments of the volume of the model, volume-to-surface ratio, or the Fourier transform of the volume of the shape. Landmarks can also be defined three-dimensionally (Lycett et al. 2006). Other global features for 3D shape are bounding boxes, cords-based, moments-based and wavelets-based descriptors, convex-hull based indices like hull *crumpliness* (the ratio of the object surface area and the surface area of its convex hull), hull packing (the percentage of the convex hull volume not occupied by the object), and hull compactness (the ratio of the cubed surface area of the hull and the squared volume of the convex hull).

Rather than working with 3D enhancements for 2D originally designed methods, an alternative approach is based on building a grid mesh joining the vertices, edges and faces defining the shape of a polyhedral object (Georgopoulos et al. 2008*)*. The faces usually consist of triangles, quadrilaterals or other simple convex polygons. Here, a polygonal mesh acts as a 3D boundary representation (*B-rep*) describing the

archaeological object as a set of surfaces that separate the object interior from the environment and is a geometrical approximation of a curved object surface.

Polygonal meshes are usually displayed as *wire-frames*, built by specifying each edge of the physical object where two mathematically continuous smooth surfaces meet, or by connecting an object's constituent vertices using straight lines or curves. The object is projected onto the computer screen by drawing lines at the location of each edge. Since wireframe renderings are relatively simple and fast to calculate, they are often used in cases where a high screen frame rate is needed. When greater graphical detail is desired, surface textures can be added automatically after completion of the initial rendering of the wireframe. This allows the designer to quickly review changes or rotate the object to new desired views without long delays associated with more realistic rendering. The study of such polygon meshes is a large sub-field of computer graphics and geometric modeling (Bertolotto et al. 1997; Howard 2005; Movchan and Movchan 1998; Ghali 2008). Different representations of polygon meshes are used for different applications and goals.

The use of thin-plate splines can be considered as another surface method of 3D shape encoding. A thin-plate is a thin sheet of some stiff material with infinite extension. When specific control points along the plate are displaced, then the plate undergoes a deformation minimizing the total bending energy E implied by the transformation (Costa and Cesar 2001).

Neural networks can be used to build connected surfaces from edge and vertices input information. In general, such programs rely on coordinate information (*x, y, z* or length, width, height), that is to say, the spatial location of interfacial boundaries for *interpolating* a geometric mesh (see Gu and Yan 1995; Piperakis and Kumazawa 2001; Barceló 2008). The advantages of using neural network for such a task are diverse (Peng and Shamsuddin 2004):

- Neural networks with backpropagation technique are able to estimate the depth (z) of an object with higher accuracy than other methods. It also means that neural networks are able to reconstruct object from 2D image to 3D after training.
- This type of reconstruction is able to produce more points of an object or surface. Therefore, a neural network is able to reconstruct more complex object with smoother surface.
- Even with scattered or unorganized data, neural networks are able to regenerate the object when outliers are removed and the smoothness of the surface is maintained.

Nowadays, laser scanned data can be easily integrated into specific software that translates clouds of points defined by x, y, z coordinates into the polygon mesh (Petersen et al. 2006; Karasik 2008). There are computer programs like PolyWorks (www.innovmetric.com), 3DReshaper (www.3dreshaper.com), PRISM 3D (www.quantapoint.com), Geomagic (www.geomagic.com), RapidForm (www.rapidform.com), among many others. In this way, scanned points can be "automatically" connected together through some algorithm, which results in a "mesh-frame" from which a smooth surface may be extrapolated. Although there

is a huge quantity of algorithms and procedures for building the polygonal mesh from scanned 3D data, in most archaeological applications the particularities of the interpolation method are not usually took into account. It is impossible to give here an overview of current applications in archaeology. For examples of polygonal meshes as 3D models of archaeological objects and entities, the reader is referred to the annual Computer Application in Archaeology series (www.caaconference.org) or the Virtual Reality, Archaeology and Cultural Heritage Research Network (www.epoch-net.org).

Archaeological models created with polygon meshes must store different types of elements. These include vertices, edges, faces, polygons and surfaces. Here, a *vertex* is a position along with other information such as color, normal vector and texture coordinates. An *edge* is a connection between two vertices. A *face* is a closed set of edges, in which a *triangle face* has three edges, and a *quad face* has four edges. A *polygon* is a set of faces. *Surfaces* are required to group smooth regions. They are needed to group smooth parts of a mesh just as polygons group 3-sided faces.

Once adequately formatted, the mesh model can be imported into analytical programs. The user can delineate regions by identifying high and low geometrical points on the wireframe, given a user-specified sensitivity level, which can be adjusted to obtain finer or coarser region definition. It is possible then to merge and split regions to transform them into analytically meaningful surfaces.

3D meshes allow capturing the usual linear measurements (i.e., length, width, thickness), but also alternative geometric attributes that researchers traditionally have had difficulty to measure accurately on real objects. Among them, we can mention the coordinates of the center of mass (CM) and the coordinates of the center of the enclosing cube (CMEC), absolute object symmetry, a surface's area, and the average angle at which surfaces intersect (Cignoni et al. 2001; Razdan et al. 2001, 2004; Riel-Salvatore et al. 2002; Simon et al. 2002; Tsirliganis et al. 2002; Ozmen and Balcisoy 2006; Grosman et al. 2008). The use of 3D polygon meshes also makes possible new measures based on topology and global or local changes in curvature that define the shape of the original object. Volume is another parameter that can be easily calculated from the polygonal mesh, although multiple lines of inference are needed to define adequately systemic function from vessel shape. Even though the attribute "capacity" for any entity working as a container (from a vessel to a building) has been relatively neglected in archaeology, it has maximum importance. Varied methods that determine a container volume exist in the literature. Most of these techniques are relevant only when the entity analyzed is highly standardized in terms of geometry. Since the very first computer applications in archaeology, volume has been calculated using formulae derived from geometry to assess the general shape of the container. Primary examples of these methods have been discussed by Castillo et al. (1968), Ericson and Stickel (1973). Nelson (1985: pp. 312–131) estimated pottery vessels volume through the calculus method of "stacked cylinders," which envisions the vessel as divided horizontally into a series of regularly sized slices. Such slices represent the diameters of very thin cylinders. Stacked one on top of another, the sum of these cylinders represents the entire vessel's volume.

In general, all those approaches can be classed among the very first geometrically based shape analysis in archaeology although these methods can generally render a useful value for a container capacity; they are usually not very accurate if the measured entity is irregular in shape and little standardized. Preliminary essays of measuring volume capacity by computer were those by M. Smith (1983, 1985), who estimated a pot capacity by integration as the volume of the solid of revolution formed by revolving the profile curve around the *x*-axis (see also Senior and Birnie 1995). Today we use the analytical possibilities of geometrical analysis software for taking such measurements. However, we should take into account that a polygonal mesh is a *boundary* representation, defining the solid in terms of its "external" surface. In the case of computationally measured volume capacity, care must be taken to compensate for real world wall thickness, then, because the volume could be overestimated, even if the geometry is well suited to the container shape.

Mara and Sablatnig (2005b) suggest measuring the curvature of a surface based on the geodesic distances among points in the neighborhood of some particular vertex of concave or convex region. Simon et al. (2002) show how once a 3D mesh has been calculated a 2D curvature mode can be estimated to quantify and measure various properties of the *B*-spline curve such as the number of inflection points, and the symmetry of the curve. Positive to negative crossings of the horizontal axis would correspond to inflection points in the profile: changes from convexity to concavity or vise versa. A *curvature plot* corresponds to the complete vessel profile.

The use of polygonal meshes to represent 3D solids allows the objective determination of the best fitting axis of rotation from the 3D data and many profile lines can be extracted, using sections through the best symmetry axes. The large number of profiles enables the archaeologist to study the uniformity of the vessel by performing a correlation analysis based on the curvature function of each profile. The correlation analysis provides a quantitative measure for the vessel uniformity, which has direct bearing on the production technology and its development in antiquity or prehistory (Simon et al. 2002; Mara et al. 2004). In the same way, it is now possible to deduce the deformations of wheel-produced pottery. A quantitative measure of the deformations can be obtained by using the polar representation of the curves that are the boundaries of the horizontal sections. About this subject, in the case of pottery see archaeological applications in Saragusti et al. (2005), Karasik (2008), in the case of lithics see: Nowell et al. (2003), Bird et al. (2007); in the case of coins see: Zaharieva et al. (2008). Doi and Sato (2005) suggest using morphing or warping techniques to study how a model can be deformed into another model. A systematic study of these deformations may reveal the technological flaws that induced them, and might possibly be used to characterize workshops methods and production patterns.

Alternatively, the medial axis or *skeleton* representation of Blum (1973) (see also Leymarie 2003 for archaeological applications) has been promulgated as a *generic* representation for describing 3D shapes from complex solids, such as those crafted by humans. The approach has shown great potential in object recognition, in solid modeling for designing and manipulating shapes, in organizing a cloud of points into

surfaces, for volumetric mesh generation, in path planning, numerical tool machining, animation, etc. On the one hand, medial or skeleton-based representations are based on a *spatial* layout, which is *thin* and tends to lie *midway* inside or outside the object's contour or profile, with a *branching* structure related to the topology of the object, forming a "stick figure." The main reasons why Blum's shape representation approach has been viewed with great promise as a universal model for shape are:

(i) A medial representation is an intuitive one to represent elongated objects, such as anthropomorphic forms, which are built upon a real skeletal frame.

(ii) It can also encode the "blobbyness" of a shape, that is, the varying width of forms, by its radius function.

(iii) Important contour *features*, such as curvature extrema and ridges, are made explicit by the medial branch "tips".

(iv) Both the interior and exterior regions of space can be described and segmented as a function of their relative closeness to the object's outline.

(v) A *partitioning* of shape is made possible by combining width and elongation properties, where, *e.g.*, different skeleton branches relate to different object parts.

(vi) A *hierarchy of scales* is also "built-in" via this combination of spatial and time-like properties, *i.e.*, smaller features can be distinguished from larger ones and ranked accordingly.

(vii) *T*he skeleton representation is *complete*, i.e., unless one starts pruning its branch structure, all of the object is represented, from the small bumps along the boundary to the large engulfing of its main parts – completeness ensures that an exact reconstruction is always possible.

(viii) The skeleton representation addresses the issue of *dimensional reduction*, i.e., it maps the entire object and the space it occupies into a thin set and has the potential for high data compression and information concentration, a property explored in particular in the domain of Computer-Aided Design (CAD).

From Fragment to Object

Archaeological evidences are fast always incomplete: not *all* past material things have remained until today. Even more, most of those few items from the past that we can observe today are broken. The only possibility to "see what cannot be seen" is as a generalization of fragmented observable data, representing partially the view of a lost physical world reality. This can only be done by generating *simulated* data (Barceló 2008). That is to say, archaeologists need a complete "model" of the original entity the fragment comes from in order to complete damaged input data. The general idea is to use a hypothetical shape model of the thing, and to fit it to the incomplete input data to simulate what is not preserved. We can use the following kinds of knowledge:

- If all we know to simulate missing data are analogies and some other "similar" cases, then we can build a *qualitative* model. This is the case of ancient buildings. In most cases, preserved remains do not shed light on the structure of vertical walls, which therefore remain unknown. In general, the reconstruction of the shape of archaeological badly preserved ancient buildings is largely based on these types of sources:
  - Pictorial evidence from plans and photographs of the building before it became a ruin.
  - Evidence shown by contemporary buildings in other neighboring places or culturally related areas, which gives clues as to likely construction methods and spatial forms.
- When old drawings and photographs are not available, external data can be estimated from ethnographic records.

The problem in all those cases is that analogical knowledge is not being added to the model in a systematic way. That is to say, the hypothesis of the original shape is not organized in rules and facts, but selecting additional information in a subjective way, using what the illustrator wants, and not what the archaeologist really needs.

In ideal terms, we should follow the rule: "The most similar is taken for the complete simulation". The procedure is as follows: we transform perceived data as a geometric data set (shape, size, texture), and we try to interpret the visual type, assuming some dependent preference function. Once the type is decided, the closest fit is determined using different numerical techniques (Barceló 2002).

IF b (x,y,z) FITS THEORY,
        And MODEL A IS A PROJECTION OF THEORY,
    THEN b (SHAPE) DERIVES FROM MODEL A.

For instance:

    IF the *geometric model of* (x) has geometric properties $A,B,C,$
    THEN (x) is an example of MODEL ABC.
    IF (x) is an example of MODEL ABC,
        And (x) has not property $D,$
    THEN JOIN property D to the geometric model of (x).

Where JOIN is an operator implemented as a command able to add some geometric unit to those already present in a preliminary model of the partial input. As a result, some new visual features (property D) are added to the geometrical model of the original data.

To deal with uncertain knowledge, a rule may have associated with it a confidence factor or a weight. For instance:

    IF the geometric model of (x) has geometric properties A,B,C but not properties
    $D,E,$
    THEN (x) is an example of MODEL ABC (with probability 0.7).

IF the geometric model of (*x*) has geometric properties *A,B,C, D,E*,
THEN (*x*) is an example of MODEL ABC (with probability 1.0).
IF (*x*) APPROXIMATELY fits MODEL ABC,
THEN *VISUALIZE* the incomplete parts of (*x*) using *A,B,C* properties.

Archaeologists need to build the model first, and then use it for simulating the unseen object. They create a geometric model of the interpreted reality, and then they use information deduced from the model when available visual data fit the model. In most cases, they create "theoretical" or "simulated" shape models. Here "theory" means general knowledge about the most probable shape of the object to be simulated or prior knowledge of the reality to be simulated. An example of this approach is provided by De Napoli et al. (2001). The reconstructed surface of each fragment is compared with the surface of a series of virtual vessels, formerly modeled, until its position is well determined on it. This operation of matching is repeated for every piece. Since vessels are typically surfaces of revolution, every typology of vessel is stored with a table of significant values of Gaussian curvature measured on a meridian with a suitable pitch value (about ten values for each item seems to be significant). Comparing the estimated datum with the stored one, the fragment can be matched to the proper parallel.

The approach by Maiza and Gaildrat (2005) is based on the exploration of a solution space constituted by all the positions that a fragment can take, relatively to a tested shape model. To achieve this exploration, it is needed a technique to evaluate the distance between a particular position of the fragment and the tested model. Genetic algorithms can be used in order to determine an optimum position of the fragment. This search is based on the previous computation of the distance to evaluate the quality of the fragment's matching. Melero et al. (2003) have also developed a semi-automatic system that uses genetic algorithms to carry out classification of potteries using rim-fragments, by mimicking the method of the archaeologists (orientation, diameter estimation, profile extraction, drawing of the fragment, additional measurements). The use of genetic algorithms permits a flexible approach adapted to the noise produced by the digitalization of the objects. A genetic algorithm based approach allows seeking multiple positions of a fragment in order to minimize the distance criterion between this fragment and an object model.

As these examples suggest in the case of regular and symmetric objects like pottery vases or containers, reconstructions are apparently easier, because we assume the pottery vessel has been created on a potter's wheel and it is hence symmetric about the *Y*-axis, fitting better with a single geometric model. That means that a reconstruction can only be performed if a priori knowledge on the type and class of vessel of which the fragment is a part is provided (Kampel and Sablatnig 2003b; Mara and Sablatnig 2008).

The reconstruction of a given object seems to be most of the times a direct generalization of fragmented observable data by mathematical object description. The fragmented spatial information available can be extrapolated to complete a closed surface. The procedure may be illustrated by the mathematical ovoid and the eggshell compared. The eggshell is a solid formed by a fine closed surface. Continuity and dynamics are bound to the shape of the eggshell, in such a way

that it is possible to locate the fragments of a broken eggshell as well as to define the whole by only very few spatial measurements. It would seem that to model the geometry of an eggshell, it would be sufficient to pick from the fragments of a broken eggshell some spatial world data to simulate the entire eggshell. The spatial continuity and dynamics of the ovoid is included in the mathematical description, to simulate the missing information.

The initial hypothesis is that the ceramic pot to which the fragment belongs was inscribed in a revolution solid. The correspondent revolution axis and the generator profile should be defined through a sequence of operations based on the geometric properties of a revolution solid. The oldest and well-approved approach is the manual method used by archaeologists for several decades. This manual approach is based on the knowledge about the production process of ceramics, manufactured on rotational plates for thousands of years. Consequently, the rotational axis is estimated orthogonal to the container orifice plane. In the same way, any horizontal section of the symmetrically produced pottery container is a circle with centre contained in the rotation axis. If the preserved rim fragment is sufficiently complete, the archaeologist can select three points on what she/he considers as defining the rim plane or any other plane parallel to the rim. This plane is used as a first approximation for the determination of the axis of symmetry of the fragment (Yao and Shao 2003). In this way, by placing the fragment with some rim vestiges on a planar plate such that the contact between the rim and the plate is maximal, the plate will then define the tangent plane, which is parallel to the plane of the original wheel. The vessel axis of symmetry is perpendicular to the tangent plane, and goes through the center of the arc generated by the contact points of the rim with the table. If we intersect the vessel with any other plane which is parallel to the tangent plane, we would find two concentric circles (or arcs when we deal with fragments), and their common center lies on the axis of rotation. Thus, by cutting the 3D representation of the vessel by several parallel planes, we can identify the axis of rotation as the line going through the centers of the concentric circles. The tangent to the rim is just a special case of this family of planes where the two circles, corresponding to the inner and outer surfaces, coalesce to a single circle.

A broken fragment reconstruction algorithm based on the above understanding, will consist of two independent steps: the first, endeavors to find the axis of rotation as the line of centers of the circular arcs, which result from the planar intersections (the horizontal sections method). The second makes use of the points of the rim, and attempts to find the best fitting plane tangent to the rim (the rim-tangent method). The horizontal sections method takes advantage of the entire information on the surface and therefore it is usually more stable and reliable. The rim-tangent method is used for fine-tuning, and it improves the quality factor in some cases.

Nevertheless, it is in no way obvious how to determine from an isolated fragment the rotational axis and the profile of the pot, even though we assume the original object was symmetric. We would need to fix the position and the orientation of each fragment so as to put the base of the object it comes from on the (x, y) plane and to align the original axis of rotation with some relevant z axis. For the fragments, the major problem is to specify their orientation respective to the z-axis. This is a non trivial problem because a great number of fragments have no determining

characteristics (a piece of rim or base, presence of external or interior grooves due to handling during the molding of the object before cooking, a more important quantity of material in one of the ends of the object).

In the two-dimensional (2D) case, where the complementary matching is reduced to a "jigsaw puzzle", many solutions have been proposed in terms of matching planar curve segments (Zhou et al. 2009). Current approaches of matching can be divided into two types:

1. Matching the fragments based on region: two sherds have common surface, and the adjacency relation between two fragments is decided my matching the shape of surfaces. The ideal situation is that two surfaces are able to match.
2. Re-assembling the fragments based on contour lines: two adjacency sherds have the same or similar edges; the object is restored by matching and aligning between each other.

Papaioannou et al. (2002) presented a method designed for 3D complementary matching of archaeological fragments, in the form of arbitrary polygonal surfaces. The method introduces a matching error between complementary surfaces that exploits the *z*-buffer algorithm. For each pair of surfaces the algorithm employs a stochastic search technique to minimize this matching error and derive the transformation that aligns the two fragments. The pair wise matching errors are used in an optimization scheme to cluster the fragments into reconstructed objects. The method uses only surface information and does not take advantage of possible boundary curve similarities. If a large percentage of the fragments are expected to participate in valid combinations, an approach that minimizes the sum of the matching errors of the individual combinations is adopted. This approach generates more fair solutions but diminishes the importance of perfect matches. More specifically, the set of fragment combinations that yields the smallest cumulative error is determined using exhaustive search.

Nevertheless, not always we can estimate the complete profile to which a fragment belongs using only information contained in the shape of the fragment. The problem is that usually the fragments usually cover a rather small part of the full perimeter of the original vessel. The smaller the fragment, the harder it is to establish its correct positioning. Furthermore, common geometric features of the fragmented objects, assuming a moderate surface deterioration, are the irregularity of the broken surfaces and the sharp curvature transition from an intact surface to a broken one (Papaioannou and Karabassi 2003). When a human tries to determine a possible match between two solids, the correspondence between the boundary lines of the surfaces is difficult to establish, because it is not always clear what faces of the polygonal mesh belong to the original surface and which ones to the fracture.

Although there are photometric pairwise approaches (Sagiroglu and Erçil 2005; Boon et al. 2008; Zhou et al. 2009) based on the estimation of the photographic affinity between the visually perceived texture of neighboring fragments, existing techniques for solving the fragment reconstruction problem mainly focus on the analysis of the break curve. Kong and Kimia (2001) were among the first to propose

an algorithmic solution for reassembling broken two-dimensional fragments comparing the curvature-encoded fragment outlines. The method was based on local shape matching followed by global search and reconstruction in the case of flat objects. Leitao et al. (2001, 2002) have proposed an alternative approach based on multiscale matching and constrained dynamic programming. Geometric pair wise matching is based on the estimating the original curvature of the complete profile and has been proposed by Papaioannou et al. (2002). They used a global optimization method to minimize an error measurement of the complementary matching between two object parts at a given relative pose, based on a point-by-point distance between the mutually visible faces of the objects. Huang et al. (2006) followed a feature-based approach in combination with a non-penetrating iterative closest point algorithm. Among other pair wise matching approaches for fragments of pottery surfaces based on the previous estimation of the axis/profile curves there is the one proposed by Cooper et al. (2002), Willis et al. (2003), Willis and Cooper (2004). They have developed a method for fragment matching based on a Bayesian approach using break curves, estimated axes and profile curves. The procedure estimates the axis/profile curve for a sherd by finding the axially symmetric algebraic surface which best fits the measured set of dense 3D points and associated normals, not requiring any local surface computations such as differentiation. Kampel and Sablatnig (2002, 2003a) proposed a matching algorithm based on the point-by-point distance between facing outlines. As the orientation of the candidate fragments was already known, the alignment of two fragments could be achieved in a two-degrees-of-freedom search space. Zhu et al. (2006) follow a partial curve matching method to find candidate matching fragment pairs. The fragment contours are represented by their turning functions and the matching segments are found by analyzing the difference curve between two turning functions directly. The curve similarity is evaluated as the residual distance of corresponding points after optimal transformation between two matching segments (see also Winkelbach et al. 2008).

When the original orientation of the pottery fragment within the complete object is not known beforehand, we may use of the fact that in wheel-produced ceramic, any horizontal section of a broken fragment is a circular arc. Thus, it will be possible an approach based on the examination of the vessel horizontal sections defined by circular arcs whose centers provide an improved estimate of the axis. This algorithm is very efficient for surfaces defined in terms of parts of cylinders or cones, that is, for surfaces whose profiles consist of straight lines. However, for surfaces formed by the rotation of a more complex profile (e.g., archaeological potsherds) this method fails (Halir 1999).

Kampel et al. (2005, 2006a) suggest finding the original orientation from the identifications of circular rills on the surface of the fragments. These rills are artifacts created from tools or fingers during the manufacturing process on the rotational plate. The first step of this method is to identify the inner side of the sherd where the rills are located. This can be done by measuring the curvature of the surface. This algorithm is based on the manual approach, where the sherd is tilted and rotated, so that the concentric rills can be seen as parallel lines, which are orientated horizontally. The authors estimate the center of gravity and the balancing plane of the

remaining vertices of the reduced inner surface. The balancing plane is described by the two longest eigen-vectors of the mean-normalized vertices, which are estimated by using the singular value decomposition. In a third step a line is fitted by minimizing the least square error to the centers of the concentric circles with the minimum variance. The fitted line is used as estimated rotational axis. It is tilted orthogonal towards the balance plane to find the best line fit. After each tilt, the centers of the concentric circles are estimated. For these centers, different lines are fitted. The line with the best fit is chosen as the rotational axis (see also Pires et al. 2007).

The vectors of unit length perpendicular to a surface are referred to as the normal vectors. Vectors normal to the fracture surfaces or to other features violating the assumed axial symmetry are not perpendicular to a surface of revolution, nor do they intersect at the axis of symmetry. Thus, the line minimizing the squared distances to the normal vectors defines the axis of symmetry. This fact can be used effectively to eliminate the irrelevant parts of the scanned surface. Grosman et al. (2008) (see also Karasik and Smilansky 2008) have suggested studying the distribution of perpendicular vector normals on the surface of the object to determine symmetry planes which estimate the object position in 3D space in terms of surface tensor calculations. Although very effective in many cases, Mara and Sablatnig (2005b) and Kampel et al. (2006b) consider that the method fails for S-shaped objects and for irregularly shaped objects.

Recently, some authors have critized the idea of a completely algorithmic approach to the reconstruction problem (Goel et al. 2005; Reuter et al. 2007; Lu et al. 2007; Karasik and Smilansky 2008). Even though automatic methods assist the assembly task by classifying and matching the fragments, they cannot fully replace a manual user interaction. Archeologists reason not only bottom up by pairwise matching, but also top-down, by considering the assembly problem as a whole, and by taking into account the archeological context. Algorithmic solutions should always be integrated, either before the manual assembly for classification and matching, or after the assembly for precise alignment.

In this context, Goel et al. (2005) propose specific software enabling the alteration of pot profiles via the addition of user-defined splines along their length. A spline can be created by altering the contour control points in number and in location. A previously saved spline can also be loaded from the database. Splines can be appended to any point along the boundary of an existing pot profile and corresponding changes to the profile and the 3D model of the pot can be viewed simultaneously. Archaeologists can utilize this feature for altering profiles of complete pots and for extending profiles of fragments. The classification of partial and complete pot profiles is done with respect to a user-defined database of profiles. To add a new profile to the database, an image of the profile is provided to the tool for computations. The profile is then added to the database, becoming available for comparisons. There also exists a separate database of user-defined splines. After creation of a spline, in the pot profile editor, it can be added into this database. This enables the archaeologist to apply the same base to several different pot profiles without having to redraw the spline.

Lu et al. (2007) use boundary curves transformed into curvature and torsion form to allow the user to view the geometrical representation of the broken fragment. The archaeologist selects one boundary curve of interest from other fragment and finds out which other fragments have a high probability of originating from the same artifact based on both the automatic ranking and her own expertise.

Too Many Different Shape Encoding. Which is the best?

It is easy to understand that there is not a single way to encode the shape of archaeological evidences, in the same way that there is not a single way of encoding the shape of any solid. As we suggested at the beginning of the paper, "shapes" do not exist in the real world as particular entities. They are instead, a particular way of seeing the world. There are not round, curve, linear things in the archaeological record, but some geometric properties of the observed entity can be described using a particular list of indexes values, a non-linear polynomial (a curve), a particular set of equations (a surface), or a topologically complex mesh (a 3D grid). There are different ways of encoding different levels of geometric information, depending on the nature of the archaeological problem we want to solve.

Conventional approaches assume that a comparatively small number of discrete features based on lengths and angles may describe the shape of any object. Thus, any archaeological material can be represented as a vector in a high-dimensional space where most dimensions are constituted of spatial dimensions with a Euclidean metric, while some other dimensions represent the angles between geometric primitives (in polar coordinates). Nevertheless, for the most part such features only capture single dimensions of variability throughout the analyzed specimens. Morphological variability is not necessarily expressed as single lines, nor do arbitrarily defined vectors constitute the only places where morphological variability is expressed.

The problem with this feature-based approach of shape is that its measurement schemes have been devised, with few exceptions in a more or less *ad hoc* manner based on what the researcher views as important dimensions of the artifact. Given that dozens of size parameters are possible, they can be combined in hundreds of ways into a formally dimensionless expression that might be used as a shape descriptor. In fact, only a few relatively common combinations are possible, but even these are plagued by inconsistency in naming and calculation conventions. The burden placed on the user is to be sure that the meaning of any particular shape descriptor is clearly understood and that it is selected because it bears some relationship to the observed changes in the shape of features, because it is presumably being measured in order to facilitate or quantify some comparison. It is important to take into account that an unlimited number of visually quite different shapes can be created with identical values for any of these dimensionless shape parameters.

Global shape encodings using all 4D information (length, width, height, and texture) can be useful in some cases, but in some other cases, they can be imprecise. Although they deliver a polygonal mesh based on a high density of scanned points,

the accuracy of individual points barely reaches 1 cm (cf. Koch 2009). On the other hand, such models usually contain too much redundant information. Of course, it can be very useful working with the full geometry of the archaeological evidences, but we do not necessary need all that information, because in most cases, we do not need all the sensorial possible visual information to solve the archaeological problem. For instance, the profile of a symmetric object can be enough for understanding the way the object was manufactured and used in the past. Even in the case of asymmetric archaeological solid materials, an analysis of the geometric irregularities along the profile or contour can allow the analysis. An excess of 4D information can exaggerate shape variability in our data. It is important to take into account that archaeological samples have been formed along periods of time comparatively long. We do not have access to all instruments made by a single artisan during a year, but a series of instruments, supposedly equivalents, made by many different artisans along three or four generations (100 years). If we are looking for regularities in shape, and interpreting measured regularity in terms of social activities performed in the past, we need to go beyond individual variability. An excess of redundancy can prevent such an analysis.

Although impressive results can be achieved for these shape encoding mechanisms, there is still the lingering problem that current automatic simplification algorithms perform poorly on a large class of manufactured objects. In contrast to current 4D visualization formats, which are dominated by smooth differential surfaces, human made objects are usually dominated by discrete features. Archaeological evidences are for the most part artificial solids containing many sharp edges and for large parts of them, a polygonal mesh is not an approximation of a smooth differential surface. Instead, the mesh represents the actual piecewise linear surface (Jang et al. 2006). If we apply current smooth simplification methods to this kind of objects, resulting simplifications may deviate from the ideal.

– Small features are merged into new larger ones. The larger features have characteristics not present in the smaller features. In this case, new face orientations are introduced.
– Many intermediate steps of the calculated simplifications are not correct
– It is not clear which intermediate simplification steps are meaningful.

Another important criticism to approaches based on interpolated global shape models is the fact they are for the most dependent on the orientation of the object, especially in the case of uneven and asymmetric objects (lithic tools, animal carcasses, material accumulations on the ground, complex buildings, etc.). Each orientation may produce a different representation, which clearly indicates that the shape of the object cannot be uniquely derived from its surface or curvature alone.

Current shape modeling systems traditionally based on polygonal meshes and other boundary representation formats presently lack the construction history and constructive object structure. From an archaeological and culture history point of view, internal structures (revealing the logic of construction and the properties of

solid material) of objects, as well as their time, and other parametric dependencies also should also be added to the consideration (Boss et al. 2009).

Therefore, shape-encoding procedures that make emphasis in the main aspects of the social processes that generated perceived shape are more convenient, than visually rich but excessively complex encodings. Of course, there is a balance between complexity of the encoding format and quantity of necessary geometric information. Sometimes harmonic analysis (Fourier decomposition or Hough transforms) can simplify excessively available information. Geometric Morphometrics based on functional landmarks (especially in the 3D case) can produce then better results than most alternative approaches.

Nevertheless, nothing in shape analysis is simple. As we will see in the next sections of this chapter, the analysis of visual data in archaeology should be based on a more advanced assumptions that the mere encoding of perceived visual regularities.

Intermediate–Level Visual Analysis

Archaeologists do not define what they "perceive" as an unorganized set of discrete, irregular, discontinuous geometrical boundaries or edges, but in terms of complex objects. That is to say, archaeological evidences are a complex aggregate of individual elements interacting with different formative/modificative agents in a statistical way. We need to distinguish between the shape of an artifact (by virtue of it being a material object) and those aspects of the material shape that arose through deliberate modification of the artifact as a whole – or in part- during its production or use.

It is then necessary to merge geometric shape encoding into larger, more complex explanatory entities. If archaeologists would simply have encoded the geometry or size of each observed entity completely separately in a spatially invariant fashion, and then tried to analyze archaeological observable evidences on the basis of the resulting collection of features, they would have lose track of the causal arrangement of these features relative to each other. We would have confused archaeological evidences that had the same features but in different arrangements. In other words, an object's shape is a function of the visible arrangements of its component parts relative to each other. Therefore, in identifying an archaeological artifact or object, or when trying to determine its functional relevance one must encode not only single unitary geometries but also the relationships between all shape components.

The solution is not to give a mere enumerative list of individual shape features or global measurements but a composite model *supervenient* on its individual parts. The hypothesis that parts have their own meaning and function to understand the complexities of a distinct shape has led many researchers to assume that shape can be specified in terms of sets of parts; the idea of an alphabet of shape minimal units is then a prevalent one. This assumption implies both a decomposition approach and a constructive framework, which describe the shape of an object as a combination of primitive elements. This different approach is based on the idea that the shape of an object has an important aspect that cannot be captured by the description of its edges and interfacial boundaries: the fact that most complex objects are perceived

as being composed of distinct parts. As we will use the term, a *part* is a restricted portion of an object that has semiautonomous status in visual perception. In other words, it is any portion detectable by a set of rules or procedures (Jang et al. 2006). Thus, shape features can be arbitrarily complex. For example, parts of a figurine's face, such as the eyes, nose, and mouth, can be considered as *parts* and modeled separately. In another case, a chair has four legs, a set and a back. In addition to such decomposition in parts, object perceptions should include the spatial relations among them (Palmer 1999, p. 348). The part decomposition approach assumes that each object can be decomposed into a small set of generic components that combine to form units depending on the relationships between the components. However, it should be emphasized that *parts* need not have a natural meaning to us (such as the nose or the eye of a prehistoric cultic figurine). Shape can be defined as a group of geometric units satisfying certain mathematical properties and coinciding with what we know of the fabrication process that produced the figurine. In addition, these *parts* do not have to be composed from disjoint groups of variables; a variable can be re-used in a multiple *parts* composite ensemble. A parts-based approach selects each *part* to represent a small group of variables that are known to be statistically dependent. Such an approach avoids devoting representational resources to weak relationships and instead allocates richer models to the stronger relationships (Schneidermann and Kanade 2004).

Consider the decomposition schema for a prehistoric cultic figurine; it has several parts as a head, a neck, a body, two arms, 2 legs. Each part has a characteristic shape and size: the head may be geometrically represented as a small triangular block, the neck a short cylinder, the body a large rectangular block, and the arms and legs long and slim cylinders. The parts are arranged in more or less specific locations. The head is attached to one end of the neck, and its other end is attached to the body. The legs are attached to the bottom of the body to support it. All such information must be represented in order to recognize some visual input as an instance of a figurine. In such a decomposition-based approach, the input variables are grouped into sets. Furthermore, the relationships within each set are more accurately modeled than those across.

Shape analysis of archaeological evidences should be based on the decomposition of solid object shapes into discrete parts, followed by the identification of those parts and their spatial and temporal relationships. We have not to forget, however, that the relationship between parts (their configuration) is equally important. It is not just that the spatial arrangements of visual features are important. The intrinsic or extrinsic features of other parts may influence the internal descriptions of the parts themselves. The perception of shape will then depend critically on the part structure of distinct features. It also depends on how these various parts are related to one another in terms of their relative positions, relative orientations, relative sizes, and so forth.

The segmentation process is one of the most difficult tasks for shape decomposition. A robust segmentation is essential for shape problems that require objects to be classified or identified individually. A weak segmentation algorithm causes the eventual failure of the whole recognition or classification process. In general,

shape segmentation algorithms follow three approaches. The first group partitions an image based on abrupt changes in detected corners along a contour; for instance, the shape of some artifacts seems to have been produced as a series of connected, curved segments, formed by the intersection of different formation processes. The ancient manufacture of the object should be reconstructed from a representation of the edges and the angles through which the edges are joined together to make a corner, the complete shape can be decomposed into a sequence of nonintersecting, connected *C*- or *S*-Shaped curved line segments. Additionally, an *S*-curve can be represented as two *C*-curves joined (one in a convex direction and the other in a concave direction) that are joined smoothly with the connection point forming an inflection point rather that a corner. In this way, not only the final shape is correctly described, but also, the formation process – human labor performed in the past- can also be reconstructed.

The second category identifies the individual shape features that are similar to a set of predefined criteria. In the case of pottery, we can mention rim, wall, and base; in the case of spears and arrows, there are distinctions among base, blade, edge, point, etc. Decomposing the shape of archaeological objects has been usually reduced to a subjective procedure of identification based on personal experience assumptions. Many authors (Henton and Durand 1991; Lycett et al. 2006; Nautiyal et al. 2006; Read 2007; Kampel and Sablatnig 2007; Hörr et al. 2007; Crompton 2007) have presented objective treatments for a rule-based or criteria based morphological segmentation.

Another group of segmentation techniques is based on finding the parts of a complex shape directly (e.g. region splitting and merging). To differentiate among parts, the perceptually salient shape components are identified. In general, this process is based on predefined criteria ranging from simple measurements such as area dimensions or circularity to complex shape descriptors. In some ways, it implies the detection of *landmarks* (see previous section).

The most common 3D shape decomposition is the Generalized Cylinder (GC) approach (Binford and Levitt 2003). The key insight is that many curved shapes can be expressed as a sweep of a variable cross section along a curved axis. Issues such as self-intersection and surface singularities do arise but shapes like a coffee pot or cup are easily handled. Compound object models, called "parts" or "assemblies," are graphs of GC primitives with affixment arcs labelled by rotation in Euler angles and a translation represented in the object-centred reference frame of the part. Transformations between objects in an assembly can be parameterized, symbolic expressions, necessary to model articulation.

Biederman (1987, 1995) calls the primitive 3-D components *geons*, which is a shortened form of *geometric units*. Each *geon* corresponds to an elementary shape (e.g., a brick, a cylinder, a curved cylinder), and all shapes are represented by combinations of *geons*. Biederman defined a set of 36 qualitatively different geons by making distinctions in some variable dimensions: cross-sectional curvature, symmetry, axis curvature, and size variation. This produces a relatively small set of distinct primitive volumes from which a huge number of object representations can be constructed by putting two or more together. Because complex objects are

conceived in Biederman theory as configurations of two or more *geons* in particular spatial arrangements, they are encoded as structural descriptions that specify both the *geons* present and their spatial relationships. If geons are the alphabet of complex 3-D objects, then spatial relations among *geons* are analogous to the order of letters in words. Biederman uses structural descriptions in which 108 qualitatively different relations can be represented between two *geons*. Some of this connections concern how they are attached (e.g. SIDE-CONNECTED and TOP-CONNECTED); others concern their relational properties, such as relative size (e.g., LARGER-THAN, SMALLER-THAN). With these *geon* relations, it is logically possible to construct more than a million different two-*geon* objects. Adding a third *geon* and its relations to the other two *geons* pushes the number of combinations into the trillions.

Although *geons* are themselves volumetric entities, Biederman theory proposes they should be identified directly from image-based features such as edges and vertices. Hummel and Biederman (1992) have built a neural network system to represent how shape analysis can be performed automatically.

Edelman (Edelman 1994; Edelman and Intrator 2002a, b) has suggested giving up this compositional representation of shape by a fixed alphabet of crisp "all-or-none" explicitly tokened primitives (such as *geons*) in favor of a fuzzy, super positional coarse-coding by an open-ended set of image fragments. This alternative approach has met with considerable success in computer vision. For example, the system described by Nelson and Selinger (1998) starts by detecting contour segments, and determines whether their relative arrangement approximates that of a model object. Because none of the individual segment shapes or locations is critical to the successful description of the entire shape, this method does not suffer from the brittleness associated with the classical structural description models of recognition. Moreover, the tolerance to moderate variation in the segment shape and location data allows it to categorize novel members of familiar object classes.

One of the relatively new constructive/deconstructive shape model approaches of higher abstraction level is the function representation approach (*FRep*). It is a generalization of traditional implicit surfaces, constructive solid geometry (CSG). This representation supports a wide class of primitive objects and operations on them. In *FRep*, a 3D object is represented by a continuous function of point coordinates as $F (x, y, z) >= 0$. A point belongs to the object if the function is non-negative at that point. The function is zero on the entire surface (called usually an *implicit surface*) of the object and is negative at any point outside the object. The function can be easily parameterized to support modeling of a parametric family of objects. In a *FRep* system, an object is represented by a tree data structure reflecting the logical structure of the object construction, where leaves are arbitrary "black boxes" primitives and nodes are arbitrary operations. The following types of geometric objects can be used as primitives (leaves of the construction tree): algebraic surfaces and skeleton-based implicit surfaces, convolution surfaces, objects reconstructed from surface points and contours, polygonal shapes converted to real functions, procedural objects (such as noise), volumetric (voxel) and other objects. Many modeling operations have been formulated which are closed on the

representation. These modeling operations include set-theoretic operations, non-linear deformations and metamorphosis, and others. *FRep* also naturally supports 4D (space-time) and multidimensional modeling using functions of several variables. The main idea of visualization is to provide a mapping of such objects to a multimedia space with such coordinates as 2D/3D world space coordinates, time, color, textures and other photometric coordinates. Time-dependent shape transformations between given objects (metamorphosis) are among the typical space-time modeling operations (Pasko et al. 1995; Pasko and Pasko 2006; Vilbrandt et al. 2004; see also Maiza and Gaildrat 2005).

Artificial Intelligence research in *function from shape* uses the idea that the sequence of parts provides some indication of the object function. For example, a hammer can be defined as a T-shaped object with geometric constraints like the head is nearly perpendicular to the handle, and the handle is positioned near the center of the head. These cannot be learned without spatial relations between parts and subparts, implying that they directly relate to the affordances of an object. The most recent simulations following this approach call for constructing a generic multi-level hierarchical description of object classes in terms of meaningful shape components. Shape meaning is derived from a large set of geometric attributes and relationships between object parts. Initially, the input range data describing each object instance is segmented, each object part is labeled as one of a few possible primitives, and each group of primitive parts is tagged by a functional symbol. Connections between primitive parts and functional parts at the same level in the hierarchy are labeled as well. Then, the generic multi-level hierarchical description of object classes is built using the functionalities of a number of object instances. During classification, a search through a finite graph using a probabilistic fitness measure is performed to find the best assignment of object parts to the functional structures of each class. An object is assigned to a class providing the highest fitness value. The scheme does not require a-priori knowledge about any class (Froimovich et al. 2002; Pechuk et al. 2005).

3D shape decomposition is much more usual in the representation of ancient buildings shape. Manferdini et al. (2008) follows an approach consisting of (i) identification of single elements, (ii) naming of the elements, (iii) identification of relations between them, and (iv) definition of the volumes they subtend. All this information is stored in a database together with find's number, geo-location and other useful archaeological details. This operation requires the help and support of archaeologists and architects to recognize transitions between different elements that constitute the find and semi automatically segment it. The semantic classification of the finds is used in the archaeological database to decide whether the object is constituted by original pieces or some of them belong to other finds and should be re-located. Furthermore, the semantic classification of the finds leads to the identification of classical orders, building functions and materials as well as extra information. The semantic segmentation is done directly on the 3D geometry using a supervised classification. Additional information such as geo-location and numbering are also added to indicate a single element within the entire set of finds. Each part is connected to an instance in a knowledge base to ease the

retrieval process in a semantics-based context. The naming of each single element and of the classes in which they can be grouped is an important process that strictly depends on archaeological and architectural considerations. While the mesh is subdivided and the model segmented, the semantic structure of the find suggests the organization of single nodes and their naming. After the segmentation phase, it is possible to re-build inner subdivision surfaces, in order to define the entire volume of each single element and node. This phase is strictly dependent on the ability of archaeologists to recognize morphological elements and constructive techniques and give volumetric interpretations. The connection between the database and the single parts of an archaeological find upgrades the traditional 2D GIS to a 3D system. Digital models are georeferenced (by point, line or area belonging to necessities), so that they also can be linked to 2D systems that are generally already available.

It should be by now obvious that shape analysis of archaeological objects must be based on the decomposition of object shapes into discrete parts, followed by the identification of those parts and their spatial and temporal relationships. In any case, we have not to forget, however, that the relationship between parts (their configuration) is equally important. It is not just that the spatial arrangements of visual features are necessary. The intrinsic or extrinsic features of other parts may influence the internal descriptions of the parts themselves. The perception of shape will then depend critically not only on the part structure of objects and how its various parts are related to one another in terms of their relative positions, relative orientations, relative sizes, and so forth. An approach may be the organization of the parts and their shape relationships in an explicit graph structure, i.e., in terms of nodes and their connectivity via a network of links (Leymarie 2003).

High Level Analysis

We have already commented that current archaeological research, like most social science research seems addressed to the mere description of archaeological evidence. In this way, shape analysis appears as a passive presentation of the visual appearance of the archaeological record. Instead, we should analyze what really happened in the past, *why* or *how* archaeological visual data acquired their actual properties of size, shape, texture, composition and spatiotemporal location.

Why an archaeological artifact has a particular *shape*? A possible answer will be "because it has a distinct appearance on the sake of its proper functioning". We have already suggested archaeologists should investigate *how* perceptual properties (shape) allow finding the social *cause* (production, use, distribution) of what we have seen at the archaeological site. In other words, material elements found in archaeological contexts are assumed to be like they physically appear to be *because* they performed some particular function in the past (Beck 1980, see also Cotterell and Kamminga 1992; McGrew 1993; St. Amant 2002; Bicici and St. Amant 2003; De Ridder 2003; Hughes 2009).

The same argument that says the geometry of the external contour or the surface is functional by virtue of its role in the performance of an activity equally says that a *stylistic* feature (i.e. decoration, color, etc.) used in a ritual or ceremony is also functional. We can distinguish between utility in these two domains by considering whether it is possible to perform the same activity with the artifact and visual feature in question in an equally efficacious manner when characteristics of the feature are changed. That is, changes in the phenomenological domain (namely the shape of the tool) may affect the efficacy with which a task can be performed. Change in a stylistic feature (decoration) may also affect its use in a ritual but for different reasons. The meaning associated with the artifact with the feature in the context of the ritual may have been altered by the trait change and thereby made the modified object unsuitable for use in the ritual (Read 2007).

In all cases, the meaning of *functioning* is related to the term *using*. Wright's suggestion (Wright 1973) was that functional analysis must depend on a notion of design (intended shape). That is to say, an object has been made to do something in a particular way, and the goal it has to fulfill can only be attained when the artifact has some particular shape among many other possible shapes. According to this interpretation, the visual appearance of any archeological evidence could be explained because *it* performs some particular function. The function of each element would be distinguished from other non-functional (or "accidental") uses by the fact that the features that define the visual appearance of the object owe its existence to this particular use (Millikan 1999).

Wright's definition would be correct, however, only in the case of objects like huts or hats, or any other instrument-like things, which have been made according to a clearly defined purpose (Millikan 1999; Neander 1991; Cummins 2000; De Ridder 2003; Hughes 2009). A naïve transposition of Wright's etiological account to the archaeological domain would presuppose that artifacts have causal histories of reproduction and selection; then one could identify the proper function of an artifact with the dispositions for which the artifact was selected or the dispositions that causally contributed to its existence. According to Nagel (1961) the task of a functional explanation would be then to explain the *presence* of an item in a system, in our case, the specific geometry of the object's surfaces. The proper way to accomplish that task is by showing the geometry to be *indispensable*. In other words: by showing the object's actual shape to be a necessary condition for the proper functioning of the thing.

There is however something odd about the intuition underlying Nagel's account (see criticism in De Ridder 2003). It supposes that a functional explanation amounts to prove the indispensability of some feature. But that is certainly not an incontestable claim. Why not suppose that a functional explanation should explain *how* an item helps to realize a function? This leads to two different conceptions of what a functional explanation is: on the one hand it can be defined as an explanation that contains a function ascription (this is what Nagel does), and on the other as an explanation of the physical mechanisms that underlie a function ascription. On the former definition the appropriate explanatory question is: "Why object *O* has shape *S*?" but on the latter it is: "Why (or how) S contributes to perform function *f*?"

A function ascription asserts that the archaeological artefact's shape is good for *something*, specifically, for bringing about some particular state of affairs. Therefore, every instrumental function includes a *functional goal*: a condition which can be realized by proper use of the item. Functional goals should be considered as *propositional functions* and not simple sentences (Hughes 2009). It is also important that some of them (such as, "cutting meat in a timely manner" or "containing water effectively") can be satisfied to greater or lesser extent. Consequently, instrumental functions apply to specific *types* of objects (typically, artefact types). This may be a matter of some controversy, since some accidental functions appear to apply only to particular tokens rather than types. In fact, it seems plausible that instrumental functions *do* primarily refer to artefact types and only derivatively to tokens. One may argue that, even for novel artefacts, the proper subject of a function ascription is an artefact type, albeit a type instantiated by a single token. We can argue that archaeological evidences functional statements should provide an answer to the question "how does *S* work?" where *S* is a goal-directed system in which the material entity whose function we are interested in appears (Nagel 1961; Boorse 1976, 2002; Adams 1979; Cummins 1975, 2000, 2002; De Ridder 2003; Hughes 2009).

Defining the shape of archaeological materials in terms of functional information is not as straightforward as it might seem. The shape of archaeological objects can be the unpremeditated material consequences of accidental actions. Tools can be used for purposes not intended by their designers and conversely, an object can be used as a tool even if it was not shaped as a tool initially (St. Amant 2002; Bicici and St. Amant 2003). The problem is that, although use actions seem to be *goal-directed activities* sometimes desirable ends are achieved through the incidental or even accidental use of an object, and consequently the shape of archaeological artifacts can also be *opportunistic* or even *accidental*, without any relationship with the hypothetical use.

A central part of the idea of a function is that not *all* effects of a structure count as its functions; some are simply accidental by-products. What is needed is a constraint on which of the visual features of a mechanical system are such that may count as performing a function. We need to know which "selected effects" of a geometry count as its function or functions, presumably in virtue of contributing to the particular class of activities or conditions of the containing system which count as that system's working *right*. In short, we need a way to separate the *function(s)* of a structure from its *accidental* effects (McClamrock 1993).

To be an archaeological artifact and to be designed and manufactured in the past by a craftsperson who had some proper use - economical or ideological, subsistential or ritual- in mind ought not to be the same. When we specify the function of an artifact, we are typically specifying the intention with which it was made. If the artifact was made in prehistory or ancient times for performing some function, then that function is (through the intentions of the artifact's creator) relevant to a causal explanation of its presence in the archaeological record. Nevertheless, without the mechanism of explicit design intentions at work, function does not explain archaeological presence. It is obvious that living beings can make tools

without being able to design them or think consciously about their proper functioning. Rather than there being a single sense of function, a large family of meanings exists. Of course, we cannot reject the idea of a tight relationship between form and function independently of any other observable or non-observable properties of the object, such as its physical structure, composition, contents, texture or contexts of use. Nevertheless, it should be restricted to items which are the consequence of purposeful activity (tools), and only in some restrictive circumstances in which the object is not being used in improper ways because there are not other alternatives.

Only certain aspects of the shape of an artifact seem to be relevant to the artifact being included as a member of a functionally salient category. Therefore, it is not very clever to try to recognize archaeological functionalities by just looking at the artifact's shape; since none uses anything by looking (except in the case of a rock-art symbol or a decorative motif). An agent that is interested in learning how a tool was used in the past either needs to look for the changes it can achieve in the physical world by using the tool or be aware of the rules governing the creation of those tools. This way, tools are no longer named specifically as vases, for instance, but as *a-tool-that-can-increase-my-abilities-of-containing-different-kinds-of-substances-by-using-the-governing-rule-number-X*. Archaeologists need to search for where these tools come from and what is the underlying functionality that we achieve while using them.

So objects have shapes in virtue of their contributing to the general disposition of the object. Only if the object in fact has a specific shape, it will have the disposition of doing something in some way. This is required by the earlier claim that function ascriptions imply disposition ascriptions, which in turn reduce to conditional statements about the behavior of an object. Such statements are only true if the component in question actually possesses the required disposition. To explain a disposition manifestation we can show them to be instances of more general laws, i.e. "laws governing the behavior of things generally, not just things having *d*" (Cummins 1975). The function should then be defined because of interpretation of a behavior of the item under an intended goal. The main result will be that we need to distinguish two types of functional explanation, one showing *that* a thing has a function and the other demonstrating *how* it performs this function (De Ridder 2003).

The proper function of an item can be defined therefore in terms of what the item has done in the past, or what it normally does or is disposed to do *in a specific context of use* (Hughes 2009), and according some well defined *user-plans* (Vermaas and Houkes 2003). That means that we should define explanatory functional analysis as implying that the shape of an archaeological object must be explained in terms of the role it plays in bringing something about, and in terms of user actions and user circumstances. For example, consider the shape of a pottery vase. Each has a crucial function assigned to it: the flat bottom is for standing the vase on a surface; the handle is for grasping the vase when lifting; the inside is for containing the liquid; the rim is for supporting the object against the lips when drinking. The assignment of causal interactions to the different shape features *defines* the object as a vase

(Leyton 1992, p. 163). We may argue, then, that the *use* of this object is specified in terms of the actions applied to it, e.g., standing up, lifting, etc., and in terms of the resulting actions that the cup applies back to the environment, e.g., conveying the liquid upward. All that means that we are describing the artifact in terms of five components:

(1) INPUTS: e.g., standing up, lifting, etc.
(2) OUTPUTS: e.g., conveying liquid
(3) STATES: physical characteristics of the cup, e.g., its shape
(4) FIRST CAUSAL RELATIONSHIP:
    e.g., lifting (input) acts on shape (state) → conveying liquid (output)
(5) SECOND CAUSAL RELATIONSHIP:
    e.g., lifting (input) acts on shape (state) → shape does not change (dynamics: next state).

The functional analysis of the object is understood as converting the input actions into the output actions. This approach equates function with causal links or goal-directedness rather than logical purpose. In other words, it asserts that anything can have a function regardless of how it came into existence (Henning 2005). The proper function of an item is its normal function, understood as a causal or dispositional act. According to Cummins, function ascriptions can be translated into conditional statements describing the behavioral regularities of an object (Cummins 1975). We may need to extend this analysis and allow that such conditionals be explained not only by the possession of dispositional properties but also by the possession of non-dispositional (categorical) properties.

DiManzo et al. (1989) regarded functional reasoning as the ability to integrate shape and function with the help of planning. They described the difficulty of separating the function of a tool from the plan it takes part in, since plans and tools evolve together and differentiate with time. Their simulated model is based on a hierarchy of levels that interact with each other. At the top level, they have a task and plan representation that uses semantic functional descriptors and functional experts for planning based on functionality of objects. The object representation level uses functional experts and geometric primitives to describe objects. The next level carries out function modeling by describing some basic functions in terms of geometric primitives, and the last level performs geometric reasoning by defining geometric constraints (see also Zhang et al. 2002).

Archaeological entities have shapes, but they should also have relationships between their physical/dynamic properties (shape) and the properties/abilities of their intended users. The affordances of any archaeological evidence become obvious in its use and/or formation process. Both involve establishing and exploiting constraints (between the user/producer and the artifact, the user/producer and the environment, and the artifact and the environment).

In the same way we should enhance shape information with causal and planning hypotheses, we can add to our simulation information about dynamics. These models will use the shape, kinematic and dynamic properties of an object (e.g. motion)

to recognize its functionality while the system observes the action that is performed with the object. The motion analysis results in several motion primitives and these are compared with previously known motion-to-function mappings. Optical flow measurements are used to derive motion information for different objects. The relevant motion is in object's coordinate system and its relation to the object on which it acts. This relation is important for establishing the mapping and creating a frame of reference. Thus, the motion is derived independently of the place of action.

What underlies this sense of the concept of form-function is that the concept is essentially historical in character, what should be obvious in an archaeological investigation! The proper function of an archaeological artifact is determined not by its shape but by its history! This is what O'Brien and Lyman have recently suggested in archaeology (O'Brien and Lyman 2002; Lyman et al. 2008). However, even if the concepts of reproduction and selection are applicable in the context of archaeology, they have a different meaning that found in a biological context, where the etiological conception of functions derives (Vermaas and Houkes 2003; De Ridder 2003).

The position I take here is that the focus of shape analysis in archaeology should be to study historical causation and not a mere form-function decision based on some classifier. If evolution is to be grounded in function, and function is to be grounded on the characteristics of the entity, then perceived shape must have a particular function not because we explain it, but for objective reasons that can be derived from the historical (evolutionary) role it played. I sustain the view that archaeological functional explanation is as a complex relational system that links physical structure, intention, settings, action, and use history.

That means that it is not the artifact's shape what will explain its appearance in some archaeological record, but the history of social actions having used that tool for different purposes at different circumstances. An item's function is to explain not only why the item has a distinctive shape, but also its causal disposition (Cummins 1975; 2002). Thus, perceived shape of archaeological evidences should be explained by the particular causal structure in which it is supposed to participate. The knowledge of the function of some perceived material element should reflect the causal interactions that someone has or can potentially have with needs, goals and products in the course of using such elements. On this view, an object's function reflects the actions that can be performed on it, given both its physical structure and the physical structure of the agent interacting with it. An object's physical structure and an agent's action specify an affordance jointly, constituting the immediate causes of a perceived function (Kitamura and Mizogouchi 1999, 2004; Chaigneau et al. 2004). It should be possible to build a model of function based on a description of the physical structure (shape) of its ancestors, namely certain reproduced physical dispositions. In that sense, both the artifact and its ancestors are part of a genetic reproduction history and are thus products of processes (Lyman et al. 2008). In some cases, it can be proved that the physical structure of the element is approximately similar to the physical structure of those ancestors, including the dispositions that correspond to the proper functions ascribed to the artifact. Only malformed, and

consequently malfunctioning is an exception of the principle that the genetic structure of the causal history provides partial justification for the belief that artifact *A* has the physical disposition (shape) that corresponds to the ascribed function. Obviously this approach cannot be applied in all circumstances, because it is wrong in the case of *new* objects and the introduction of novelty and revolutionary changes, but it can be useful for understanding the causal history (or "genetic" reproduction) of a historically connected series of objects (Chaigneau et al. 2004; Rovner 2006).

This approach is at odds with current practice in morphological analysis in archaeology, still based on the assumption that objects most similar to a given function will be found in the core of the definition, while objects that are less similar will be in more marginal parts (see discussion in Barceló 2008). What we need is a study of covariances between shape representations and other associated or causal variables (human labor). In other words, whereas shape cannot be reduced to the identification of objects per se, it should be concerned with the degree to which other variables (e.g., time, space, composition, ecology, phylogeny, and use constraints) are related to shape variation within a sample or population and the nature of that relationship. As such, morphometric data analysis (Lele and Richtsmeier 2001) can be used to address a far wider range of shape-related problems than either geometry or pattern recognition. The hypothesis under examination: "is whether the social or historical events (in the physical sense of the word) cause or covary with the aspects of the objects' morphological variation (= the aspect(s) of their geometry) actually measured or extractable from the data under analysis". Confusion over this issue inevitably leads to erroneous interpretations of morphometric results (see MacLeod 2002).

If this approach were right, then to be able to ascribe functions to observed archaeological evidences, we would need to combine three kinds of information:

- Knowledge about how the designers intended to design the artifact to have the function
- Knowledge about how the designers determined the physical structure of that artifact on the basis of their technological abilities
- Knowledge about how the artifact was determined by its physical structure to perform that function

References

Adams, F.R., 1979, A goal-state theory of function attributions, *Canadian Journal of Philosophy*, 9: 493–518.

Adams, D.C.F., Rohlf, J., Slice, D.E., 2004, Geometric morphometrics: Ten years of progress following the 'revolution'. *Italian Journal of Zoology*, 71: 5–16.

Adán, M., Barcelo, J.A., Pijoan-Lopez, J.A., Pique, R., Toselli, A., 2003, Spatial Statistics in Archaeological Texture Analysis. In *The Digital Heritage of Archaeology*. Edited by M. Doerr and A. Sarris. Published by the Archive of Monuments and Publications. Hellenic Ministry of Culture. Athens, Greece.

Andrefsky, W., 2005, *Lithics: Macroscopic Approaches to Analysis*. Cambridge University Press, Cambridge.

Andrefsky, W., 2006, Experimental and archaeological verification of an index of retouch for hafted bifaces. *American Antiquity*, 71(4): 743–757.

Arlinghaus, S., 1994, *Practical Handbook of Curve Fitting*. London, CRC Press.

Avern, G. (2010, in press). "Shortcomings of Current 3D Data Acquisition Technologies for Graphical Recording of Archaeological Excavations." Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 32nd Conference, Prato, Italy, April 2004.

Barceló, J.A., 2000, Visualizing what might be. An Introduction to Virtual Reality in Archaeology. In *Virtual Reality in Archaeology*. Edited by J.A. Barcelo, M. Forte and d. Sanders. ArcheoPress, Oxford. British Archaeological Reports (S843), pp. 9–36. (http://prehistoria.uab.cat/Barcelo/publication/VR2000.pdf).

Barceló, J.A., 2001, Virtual reality and scientific visualization. working with models and hypothesis international. *Journal of Modern Physics C*, 12(4): 569–580.

Barceló, J.A., 2002, Virtual Archaeology and Artificial Intelligence. In *Virtual Archaeology*. Edited by F. Nicolucci. ArchaeoPress, Oxford. BAR International Series S1075, pp. 21–28. (http://prehistoria.uab.cat/Barcelo/publication/VAST_021.pdf).

Barceló, J.A., 2005, A Science Fiction Tale? A robot called archaeologist. In *The World is in your Eyes. Proceedings of the XXXIII Computer Applications and Quantitative Methods in Archaeology Conference*. Edited by A. Figueiredo and G. Velho. Tomar, Portugal, pp. 221–230. Associaçâo para o Desenvolvimento das Aplicações Informáticas e Novas Tecnologias em Arqueologia. (http://prehistoria.uab.cat/Barcelo/publication/SciFiTomar.pdf).

Barceló, J.A., 2007, Automatic Archaeology: Bridging the gap between Virtual Reality, Artificial Intelligence, and Archaeology. In *Theorizing Digital Cultural Heritage. A critical Discourse*. Edited by F. Cameron and S. Kenderdine. The MIT Press, Cambridge, MA, pp. 437–456.

Barceló, J.A., 2008, *Computational Intelligence in Archaeology*. Henshey (NY). Information Science Reference (the IGI Group.Inc.).

Barceló, J.A. Pijoan-Lopez, J., 2004, Cutting or Scrapping? Using neural Networks to Distinguish Kinematics in Use Wear Analysis In Enter the Past. The E-way into the Four Dimensions of Culture Heritage . Edited by Magistrat der Stadt Wien. ArcheoPress, Oxford, BAR International Series 1227, pp. 427–431. (http://prehistoria.uab.cat/Barcelo/publication/CuttingScrapping.pdf).

Barceló, J.A., De Castro, O., Travet, D., Vicente, O., 2003, A 3d Model of an Archaeological Excavation. In *The Digital Heritage of Archaeology. Computer Applications and Quantitative methods in Archaeology*. Edited by M. Doerr and A. Sarris. Hellenic Ministry of Culture. Archive of Monuments and Publications. (http://prehistoria.uab.cat/Barcelo/publication/3Dmodel.pdf).

Barceló, J.A, Mameli, L., Maximiano, A., Vicente, O., 2009, New computational and mathematical methods for archaeological fieldwork at the extreme south of the populated World. *Arctic Anthropology*, 46 (2):.

Barceló, J.A., Pijoan, J., Vicente, O., 2001, Image Quantification as Archaeological Description. In *Computing Archaeology for Understanding the Past*. Edited by Z. Stancic and T. Veljanovski. BAR International Series S931, pp. 69–78. (http://prehistoria.uab.cat/Barcelo/publication/CAA2000.pdf).

Bardossy, A., Schmidt, F., 2002, GIS approach to scale issues of perimeter-based shape indices for drainage basins. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, 47(6): 931–942.

Bebis, G., Papadourakis, G., Orphanoudakis, S., 1998, Recognition Using Curvature Scale Space and Artificial Neural Networks. In *Proceedings of the IASTED International Conference Signal and Image Processing* October 27–31, 1998, Las Vegas, NV.

Beck, B.B. 1980, *Animal Tool Behavior: The Use and Manufacture of Tools*. New York, Garland Press.

Bello, S., Soligo, C., 2008, A new method for the quantitative analysis of cutmark micromorphology. *Journal of Archaeological Science* 35, pp. 1542–1552.

Belongie, S., Malik, J., Puzicha, J., 2002, Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(24): 509–522.

Bernardini, F., Rushmeier, H., 2002, The 3D model acquisition pipeline. *Computer Graphics Forum*, 21(2): 149–172.

Bertolotto, M., Bruzzone, E., De Floriani, L., 1997, Geometric Modelling and Spatial Reasoning. In *Artificial Vision. Image Description, Recognition and Communication*. Edited by V. Cantoni, S. Levialdi and V. Roberto, New York, Academic Press, pp. 107–134.

Bicici, E., St. Amant, R., 2003, *Reasoning about the Functionality of Tools and Physical Artifacts*. Technical Report TR-2003-22, Department of Computer Science, North Carolina State University, April, 2003.

Biederman, I., 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2): 115–147.

Biederman, I., 1995, Visual Object Recognition. In *An Invitation to Cognitive Science*, 2nd edition, Volume 2., *Visual Cognition*. Edited by S.F. Kosslyn and D.N. Osherson. Chapter 4, MIT Press, Cambridge, MA, pp. 121–165.

Binford, T.O., Levitt, T.S., 2003, Evidential reasoning for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(7): 837–851.

Bird, C., Minichillo, T., Marean, C.W., 2007, Edge damage distribution at the assemblage level on middle stone age lithics: An image-based GIS approach. *Journal of Archaeological Science* 34(5): 771–780.

Bishop, G., Cha, S., Tappert, C.C., 2005, Identification of Pottery Shapes and Schools Using Image Retrieval Techniques. *Proceedings of MCSCE, CISST,* Las Vegas, NV, (http://csis.pace.edu/~ctappert/srd2005/c2.pdf). Downloaded on April 2009.

Bisson, M.S., 2000, Nineteenth century tools for twenty first century archaeology? Why the Middle Paleolithic typology of François Bordes must be replaced. *Journal of Archaeological Method and Theory*, 7: 1–48.

Blum, H., 1973, Biological shape and visual science. *Journal of Theoretical Biology*, 38: 205–287.

Bookstein, F.L., 1991, *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

Boon, P.J., Pont, S.G., van Oortmerssen, M., 2008, Acquisition and Visualization of Cross Section Surface Characteristics for Identification of Archaeological Ceramics. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Berlin, Germany, April 2–6, 2007. Edited by A. Posluschny, K. Lambers and I. Herzog, Kolloquien zur Vor- und Frühgeschichte, Vol. 10. Bonn Dr. Rudolf Habelt GmbH.

Boorse, C., 1976, Wright on functions. *Philosphical Review*, 85: 70–86.

Boorse, C., 2002, A Rebuttal on Functions. In *Functions. New Essays in the Philosophy of Psychology and Biology*. Edited by A. Ariew, R. Cummins and M. Perlman. Oxford University Press, Oxford.

Boss, M.A., Meister, M., Rietzel, D., 2009, Inside Greek vases-On Estimating the Skill of Ancient Greek Craftsmen Producing Complex 3D Shapes Using Current Technologies. In *Making History Interactive*. *Proceedings of the 37th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Williamsburg, VA, March 22–26, 2009. Edited by B. Frischer and L. Fisher. (in press).

Brown, C.T., 2001, The fractal dimensions of lithic reduction, *Journal of Archaeological Science*, 28: 619–31.

Brown, C.T., Witschey, W.R.T., 2003a, Fractal Fragmentation of Archaeological Ceramics. *Fractals in Archaeology Symposium*, Society of American Archaeology, Milwaukee, WI.

Brown, C.T., Witschey, W.R.T., 2003b, The fractal geometry of ancient Maya settlement. *Journal of Archaeological Science*, 30: 1619–1632.

Brown, C., Witschey, W., Liebovitch, L., 2005, The broken past: Fractals in archaeology. *Journal of Archaeological Method and Theory*, 12(1): 37–78.

Caldoni, M.-I., Chimienti, A., Nerino, R., 2006, Automatic Coarse Registration by Invariant Features. In *The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage.* VAST2006. Edited by M. Ioannidis, D. Arnold, F. Nicolucci and K. Mania. Budapest, Archeolingua.

Carbonetto, P., de Freitas, N., Barnard, K., 2004, A Statistical Model for General Contextual Object Recognition In *8th European Conference on Computer Vision,* Prague, Czech Republic, May 11–14. *Proceedings, Part I.* Edited by T. Pajdla and J. Matas. Springer Lecture Notes in Computer Science, Vol. 3021.

Cardillo, M., 2005, Explorando la variación en las morfologías líticas a partir de la técnicas de análisis de contornos. El caso de las puntas de proyectil del holoceno medio-tardío de la Puna de Salta (San Antonio de los Cobres, Argentina). *Revista Werken* N°7. 77–88.

Cardillo, M., 2006, Temporal Trends in the Morphometric Variation of the Lithic Projectile Points During The Middle Holocene Of Southern Andes (Puna Region). A Coevolutionary Approach. In *15th Congress Of International Union For Prehistoric And Protohistoric Sciences.* Workshop 22: Theoretical And Methodological Issues In Evolutionary Archaeology: Toward An Unified Darwinian Paradigm. Lisbon, Portugal, 4–9 September 2006.

Cardillo, M., Charlin, J., 2007, Tendencias observadas en la variabilidad de los raspadores de norte y sur de Patagonia. Explorando las interrelaciones entre forma, tamaño e historia de vida. *2do Congreso Argentino y 1er Congreso Latinoamericano de Arqueometría*. Buenos Aires.

Casali, F., 2006, X-Ray and Neutron Digital radiography and Computed Tomography for Cultural heritage. In *Physical Techniques in the Study of Art, Archaeology and Cultural Heritage*. Edited by D. Bradley and D.C. Creagh. Vol. 1, Elsevier Science & Technology, Amsterdam, pp. 41–123.

Castillo Tejero, N., Litvak, J., 1968, Un Sistema de Estudio Para Formas de Vasijas. *Technologia* 2. Departamento de Prehistoria, Instituto Nacional de Antropologia e Historia, Mexico City.

Chaigneau, S.E., Barsalou, L.W. Sloman, A., 2004, Assessing the causal structure of function. *Journal of Experimental Psychology: General* , 133(4): 601–625.

Christenson, A.L., 1986, Projectile point size and projectile aerodynamics: An exploratory study. *Plains Archaeologist*, 31: 109–128.

Cignoni, P., Callieri, M., Scopigno, R., Gori, G., Risaliti, M., 2006, Beyond manual drafting: A restoration-oriented system. *Journal of Cultural Heritage*, 3(7): 1–12.

Clarkson, C., 2002, An index of invasiveness for the measurement of uni-facial and bifacial retouch: A theoretical, experimental and archaeological verification. *Journal of Archaeological Science*, 29: 65–75.

Collins, S., 2008, Experimental investigations into edge performance and its implications for stone artefact reduction modeling. *Journal of Archaeological Science*, 35: 2164–2170.

Cooper, D.B., Willis, A., Andrews, S., Baker, J., Cao, Y., Han, D., Kang, K., Kong, W., Leymarie, F.F., Orriols, X., Velipasalar, S., Vote, E.L., Joukowsky, M.S., Kimia, B.B., Laidlaw, D.H., Mumford, D., 2002, Bayesian pot-assembly from fragments as problems in perceptual-grouping and geometric-learning. *Proceedings of the 16th International Conference on Pattern Recognition*, 3: 30927–30931.

Costa, L.d.F., Cesar, R.M., Jr., 2001, *Shape Analysis and Classification. Theory and Practice.* CRC Press, London.

Cotterell, B., Kaminga, J.,, 1992, *Mechanics of Pre-industrial Technology : An Introduction to the Mechanics of Ancient and Traditional Material Culture*. Cambridge University Press, Cambridge.

Crompton, S.Y., 1995, The third measure: 3-D data, data capture systems and accuracy. *Archaeological Computing Newsletter*, 44: 5–11.

Crompton, S., 2007, 3D Lithics. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Berlin, Germany, April 2–6, 2007. Edited by A. Posluschny, K. Lambers and I. Herzog. Kolloquien zur Vor- und Frühgeschichte, Vol. 10. Bonn Dr. Rudolf Habelt GmbH.

Cummins, R., 1975, Functional analysis. *Journal of Philosophy*, 72/20: 741–765.

Cummins, R., 2000, How does it work? vs. What are the laws? Two conceptions of psychological explanation. In *Explanation and Cognition*. Edited by F. Keil and R. Wilson. The MIT Press, Cambridge, MA, pp. 117–145.

Cummins, R., 2002, Neo-Teleology. In *Functions. New Essays in the Philosophy of Psychology and Biology*. Edited by A. Ariew, R. Cummins and M. Perlman. Oxford University Press, New York.

DeBoer, W.R., 1980, Vessel shape from rim sherds: An experiment on the effect of the individual illustrator. J*ournal of Field Archaeology*, 7: 131–135.

De Napoli, L., Luchi, M.L., Muzzupappa, M., Rizzati. S., 2001, A Semi-Automatic Procedure for the Recognition and Classification of Pieces of Archaeological Artefacts, In *12th Adm International Conference*, September 5–7, Rimini, Italy, 2001.

De Ridder, J., 2003, Functional Explanation for Technical Artifacts, *6th Inland Northwest Philosophy Conference*, May 2–4, 2003, Moscow, ID/Pullman, WA.

Dibble, H.L., Chase, P.H., 1981, A new method for describing and analyzing artifact shape. *American Antiquity*, 46: 178–197.

Dibble, H.L., 1997, Platform variability and flake morphology: A comparison of experimental and archeological data and implications for interpreting prehistoric lithic technological strategies. *Lithic Technology*, 22: 150–170.

Dierckx, P., 1995, *Curve and Surface Fitting with Splines.* Oxford University Press, Oxford.

DiManzo, M., Trucco, E., Giunchiglia, F., Ricci F., 1989, FUR: Understanding functional reasoning. *International Journal of Intelligent Systems*, 4: 431–457.

Dimitrov, L.I., Wenger, E., Šrámek, M., Trinkl, E., Lang-Auinger, M., 2006, VISAGE: An Integrated Environment for Visualization and Study of Archaeological Data Generated by Industrial Computer Tomography. In *The E-volution of Information Communication Technology in Cultural Heritage. Where Hi-Tech Touches the Past: Risks and Challenges for the 21st Century*. Edited by M. Ioannides, D. Arnold, F. Niccolucci and K. Mania. EPOCH/ Archaeolingua. (http://public-repository.epoch-net.org/publications/VAST2006/short0.pdf). Downloaded on April 2009.

Djindjian, F., 1993, *Les Méthodes en Archéologie.*Armand Colin, Paris.

Doi, J., Sato, W., 2005, Surface Reconstruction and 3D Shape Processing for Cultural Applications, *IVCNZ 2005 - Image and Vision Computing*. New Zealand 28 November–29 Novmber 2005 (http://pixel.otago.ac.nz/ipapers/38.pdf). Downloaded on April 2009.

Dryden, I.L., Mardia, K.V., 1998, *Statistical Shape Analysis.*Wiley, Chichestern.

Durham, P., Lewis, P.H., Shennan, S.J., 1990, Artefact Matching and Retrieval Using the Generalised Hough Transform. In *Proceedings of Computer Applications in Archaeology*. pp. 25–30.

Edelman, S., 1994, *Representation and Recognition in Vision*. The MIT Press, Cambridge, MA.

Edelman, S, Intrator, N., 2002a, Visual Processing of Object Structure. In *The Handbook of Brain Theory and Neural Networks* (2nd ed.) Edited by M.A. Arbib, MIT Press, Cambridge, MA.

Edelman, S., Intrator, N., 2002b, Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, 27: 73–110.

Ericson, J.E., Stickel, E.G., 1973, A proposed classification system for ceramics. *World Archaeology*, 4: 357–367.

Ericson, J.E., Read, D.W., Burke, C., 1972, Research design: The relationship between the primary functions and the physical properties of ceramic vessels and their implications for ceramic distribution on an archaeological site. *Anthropology UCLA*, III(2): 84–95.

Farjas, M., García Lázaro, F.J., (eds). 2008, *Modelización tridimensional y sistemas láser escaner 3D aplicados al patrimonio histórico.*Ediciones La Ergástula, Madrid.

Fleming, B., 1999, *3d Modeling and Surfacing*. Morgan Kaufmann Publishers, Inc., San Francisco.

Flenniken, J. Raymond, A., 1986, Morphological projectile point typology: Replication, experimentation and technological analysis. *American Antiquity*, 51(3): 603–614.

Forel, B., Gabillot, M., Monna, F., Forel, S., Dommergues, C.H., Gerber, S., Petit, C., Mordant, C., Chateau, C., 2009, Morphometry of middle bronze age palstaves by discrete cosine transform. *Journal of Archaeological Science*, 36: 721–729.

Forsyth, D.A., Ponce, J., 2003, *Computer Vision: A Modern Approach.* Prentice-Hall, Upper Saddle River, NJ.

Froimovich, G., Rivlin, E., Shimshoni, I., 2002, Object Classification by Functional Parts. *Proceedings of the First Symposium on 3D Data, Processing, Visualization and Transmission*, pp. 648–655.

Georgopoulos, A., Ioannidis, C., Ioannides, M., 2008, 3D Virtual Reconstructions at the Service of Computer Assisted Archaeological Measurements. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Berlin, Germany, April 2–6, 2007. Edited by A. Posluschny, K. Lambers and I. Herzog. Kolloquien zur Vor- und Frühgeschichte, Vol. 10. Bonn Dr. Rudolf Habelt GmbH.

Gero, J., Mazullo, A., 1984, Analysis of artifact shape using Fourier series in closed form. *Journal of Field Archaeology*, 11: 315–322.

Ghali, S., 2008, *Introduction to Geometric Computing.* Springer, London.

Gilboa, A., Karasik, A., Sharon, I., Smilansky, U., 2004, Towards computerized typology and classification of ceramics. *Journal of Archaeological Science*, 31: 681–694.

Goel, S., Jain, A., Singh, P., Bagga, S., Batra, S., Gaur, U., 2005, Computer Vision Aided Pottery Classification and Reconstruction, *INDO US Science and Technology Forum on Digital Archaeology* (http://www.siddharthbatra.info/Data%20Files/CV%20Based%20Class%20&%20Recon%20of%20Pottery.pdf). Downloaded on April 2009.

Gonzalez, R.C., Woods, R.E., 2007, *Digital Image Processing* (3rd Edition) Prentice-Hall, Upper Saddle River, NJ.

Grimson, W.L., 1991, *Object Recognition by Computer: the Role of Geometric Constraints.* MIT Press, Cambridge, MA.

Grosman, L., Smikt, O., Smilansky, U., 2008, On the application of 3-D scanning technology for the documentation and typology of lithic artifacts. *Journal of Archaeological Science*, 35: 3101–3110.

Gu, P., Yan, X., 1995, Neural network approach to the reconstruction of freeform surfaces for reverse engineering. *Computer Aided Design*, 27(1): 59–64.

Hagstrum, M., Hildebrand, J., 1990, The Two-curvature method for reconstructing ceramic morphology. *American Antiquity*, 55: 388–403.

Halir, R., 1999, An Automatic Estimation of the Axis of Rotation of Fragments of Archaeological Pottery: A Multi-step Model-based Approach. In *Proceedings of the Seventh International Conference in Central Europe on Computer Graphics. Visualization and Interactive Digital Media* (WSCG'99). Edited by V. Skala.

Hall, N.S., Laflin, S., 1984, A Computer Aided Design Technique for Pottery Profiles. In: *Computer Applications in Archaeology*. Edited by S. Laflin. Computer Center, University of Birmingham, Birmingham, pp. 178– 188.

Hardaker, T., Dunn, S., 2005, The flip test – A new statistical measure for quantifying symmetry in stone tools. *Antiquity*, 79(306).

Heideman, G., 2005, The long-range saliency of edgeand corner-based salient points. *IEEE Transactions On Image Processing*, 14(11): 1701–1706.

Henning, B., 2005, "Functional Reasoning". Draft for the Conference "Norms, Reasoning and Knowledge in Technology". *IFOMIS Saarbrücken/HU Berlin.* (http://www.borishennig.de/texte/2005/fctreasonII.pdf). Consulted on July 30th, 2005.

Henton, G., Durand, S.R., 1991, Projectile point measurement and classification using digital image processing. *Journal of Quantitative Anthropology*, 3: 53–82.

Hermon, S., 2008, Reasoning In 3D: A Critical Appraisal Of The Role Of 3D Modelling And Virtual Reconstructions In Archaeology. In *Beyond Illustration: 2D and 3D Digital Technologies as Tools for Discovery in Archaeology*. Edited by B. Frischer. ArcheoPress, Oxford, pp. 36–45 (British Archaeological Reports).

Hermon, S., Petrone, M., Calori, L., 2001, An Experimental Method For The Analysis of Attributes of Flint Artefacts Using Image Processing. In *Computing Archaeology for Understanding the Past. Proceedings of the 28th Conference,* Ljubljana, Slovenia, 18–21 April 2000. Edited by Z. Stancic and T. Veljanovski. BAR. International Series 931, Archaeopress, Oxford, pp. 91–98.

Hörr, C., Brunner, D., Brunnett, G., 2007, Feature extraction on axially symmetric pottery for hierarchical classification. *Computer-Aided Design & Applications*, 4(1–4): 375–384.

Howard, W.E., 2005, *Introduction to Solid Modeling,* McGraw-Hill, New York.

Huang Q.-X., Flöry S., Gelfand N., Hofer M., Pottmann H., 2006, Reassembling fractured objects by geometric matching. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)* 25(3): 569–578.

Hughes, J., 2009, An artifact is to use: An introduction to instrumental functions. *Synthese*, 168: 179–199.

Hummel, J.E., Biederman, I., 1992, Dynamic binding in a neural network for shape-recognition. *Psychological Review*, 99(3): 480–517.

Jang, J., Wonka, P., Ribarsky, J., Shaw, C.D., 2006, Punctuated simplification of man-made objects. *Visual Computer*, 22(2): 136–145.

Jelinek, H.F., Jones, C.L., Warfel, M.D., 1998, Is there meaning in fractal analysis? *Complexity International*, 6.

Juhl, K., 1995, *The relationship between vessel form and function. A methodological study.* AmS-Skrifter 14, Akeologisk museum i Stavanger (Norway). 1–143.

Kampel, M., Melero, F.J., 2003, Virtual Vessel Reconstruction From A Fragment's Profile. In *Proceedings of the 4th VAST.* Edited by D. Arnold, A. Chalmers and F. Niccolucci, pp. 79–88.

Kampel, M., Sablatnig, R., 2002, Automated Segmentation Of Archaeological Profiles For Classification. In *International Conference on Pattern Recognition*. Edited by R. Kasturi, D. Laurendeau and C. Suen. pp. 57–60.

Kampel, M., Sablatnig, R., 2003a, Profile based Pottery Reconstruction. In: *Proceedings of IEEE/CVPR Workshop on Applications of Computer Vision in Archaeology*, CD-ROM. Edited by D. Martin. Madison, WI.

Kampel, M., Sablatnig R., 2003b, An automated pottery archival and reconstruction system. *Journal of Visualization and Computer Animation*, 14(3): 111–120.

Kampel M., Sablatnig R., 2006, 3D Data Retrieval for Archaeological Pottery. In: *Interactive Technologies and Sociotechnical Systems, Proceedings of 12th International Conference, VSMM 2006.* Edited by H.Z.Z. Pan, H. Thwaites, A.C. Addison and M. Forte. Xi′an, China, October 18–20, Springer Lecture Notes in Computer Science Vol. 4270, pp. 387–395.

Kampel M., Sablatnig R., 2007, Rule based system for archaeological pottery classification. *Pattern Recognition Letters*, 28: 740–747.

Kampel M., Sablatnig R., Mara H., 2005, Robust 3D Reconstruction of Archaeological Pottery based on Concentric Circular Rills. In *WIAMIS05: The 6th International Workshop on Image Analysis for Multimedia Interactive Services*. Edited by N. Magnenat-Thalmann and J.H. Rindel. Montreux, Switzerland, pp. 14–20.

Kampel, M., Sablatnig, R., Mara, H., 2006a, 3D Acquisition of Archaeological Fragments and Web based 3D Data Storage. In: Proceedings of CAA06: Digital Discovery: Exploring New Frontiers in Human Heritage. Edited by J.T. Clark and E.H. Hagemeister. Fargo, April 18–23.

Kampel, M., Mara, H., Sablatnig, R., 2006b, Automated Investigation of Archaeological Vessels. In *Proceedings of EUSIPCO2006: 13th European Signal Processing Conference*. Edited by M. Luise. Florence, Italy, September 4–8.

Kampffmeyer, U., Zamperoni, P., Teegen, W.R., Graça, L., 1988, *Untersuchungen zur rechn-ergestützen Klassifikation der Form von keramik*. Frankfurt a.M. (Germany). Verlag Peter Lang (Arbeiten zur Urgeschichte des menschen, band 11).

Karasik, A., 2008, Applications Of 3D Technology As A Research Tool In Archaeological Ceramic Analysis. In *Beyond Illustration: 2d and 3d Digital Technologies as Tools for Discovery in Archaeology*. Edited by B. Frischer. ArcheoPress, Oxford (British Archaeological Reports). pp. 111–124.

Karasik, A., Smilansky, U., 2008, 3D scanning technology as a standard archaeological tool for pottery analysis: Practice and theory. *Journal of Archaeological Science*, 35: 1148–1168.

Karasik, A., Mara, H., Sablatnig, R., Smilansky, U., 2005, Measuring Deformations of Wheel-Produced Ceramics using High Resolution 3D Reconstructions. In *The World is in your Eyes. Proceedings of the XXXIII Computer Applications and Quantitative Methods in Archaeology Conference*. Edited by A. Figueiredo and G. Velho. Tomar, Portugal. Associaçâo para o Desenvolvimento das Aplicações Informáticas e Novas Tecnologias em Arqueologia.

Kardjilov, K., Fiori, F., Giunta, G., Hilger, A., Rustichelli, F., Strobl, M., Banhart, J., Triolo, R., 2006, Neutron tomography for archaeological investigations. *Journal of Neutron Research*, 14(1): 29–36.

Kendall, D.G., 1977, The diffusion of shape. *Advances in applied probability*, 9: 428–430.

Kendall, D.G., Barden, D., Carne, T.K., Le, H., 1999, *Shape and Shape Theory*. Wiley, Chichester.

Kennedy, S.K., Lin, W.H., 1988, A fractal technique for the classification of projectile point shapes. *GeoArchaeology*, 3(4): 297–301.

Keogh, E., Lee, S.H., Zhu, Q., Wang, X., Rampley, T., 2009, Towards indexing and Datamining All the World's Rock-Art. In *Making History Interactive. Computer Applications and Quantitative Methods in Archeology Conference*, Willimsburg, VA, March 22–26, 2009. (in press).

Kitamura, Y., Mizogouchi, R., 1999, An Ontology of Functional Concepts of Artifacts. Artificial Intelligence Research Group. The Institute of Scientific and Industrial Research. Osaka University (AI-TR-99-1). (http://www.ei.sanken.osaka-u.ac.jp/pub/kita/kita-tr9901.pdf).

Kitamura, Y., Mizogouchi, R., 2004, Ontology-based systematization of functional knowledge. *Journal of Engineering Design*, 15(4): 327–351.

Koch, M., 2009, Combining 3D laser-Scanning and close-range Photogrammetry-An approach to Exploit the Strength of Both methods. In *Making History Interactive. Computer Applications and Quantitative Methods in Archeology Conference*, Willimsburg, VA, March 22–26, 2009. (in press).

Kong, W., Kimia, B.B., 2001, On solving 2d And 3d puzzles using curve matching. I*EEE Conference on Computer Vision and Pattern Recognition*, 2: 583.

Kuhn, S.L., 1990, A geometric index of reduction for unifacial stone tools. *Journal of Archaeological Science*, 17: 585–593.

Kulkarni, A.D., 2001, *Computer Vision and Fuzzy-neural Systems*. Prentice Hall, Upper Saddle River.

Lambers, K., Remondino, F., 2007, Optical 3D Measurement Techniques In Archaeology: Recent Developments And Applications. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Berlin, Germany, April 2–6, 2007. Edited by A. Posluschny, K. Lambers and I. Herzog, Kolloquien zur Vor- und Frühgeschichte, Vol. 10. Bonn Dr. Rudolf Habelt GmbH.

Lambers, K., Eisenbeiss, H., Sauerbier, M., Kupferschmidt, D., Gaisecker, T., Sotoodeh, S., Hanusch, T., 2007, Combining photogrammetry and laser scanning for the recording and modelling of the late intermediate period site of Pinchango Alto, Palpa, Peru. *Journal of Archaeological Science*, 34(10) : 1702–1712.

Laplace, G., 1972, *La Typologie Analytique et Structurale*. CNRS, Paris.

Le, H., Small, C.G., 1999, Multidimensional scaling of simplex shapes. *Pattern Recognition*, 32: 1601–1613.

Leese, M.N., Main, P.L., 1983, An Approach to the Assessment of Artefact Dimension as Descriptors of Shape. In *Computer Applications in Archaeology 1983*. Edited by J.G.B. Haigh. University of Bradford, School of Archaeological Sciences, Bradford, pp. 171–180.

Leitao, H.D., Da Gama, D., Stolfi, J., 2001, Digitization and Reconstruction of Archaeological Artifacts *XIV Brazilian Symposium on Computer Graphics and Image Processing* (SIBGRAPI′01) (http://csdl2.computer.org/comp/proceedings/sibgrapi/2001/1330/00/13300382.pdf).

Leitao, H.D. Da Gama, D., Stolfi, J., 2002, A Multiscale method for the reassembly of two-dimensional fragmented objects. *IEEE Transactions on Pattern Analysis and machine Intelligence*, vol. 24 (9), pp.1239-1251.

Leitao, H.D. Da Gama, D., Stolfi, J., 2005, Measuring the information content of fracture lines international. *Journal of Computer Vision*, 65(3): 163–174.

Lele, S.R., Richtsmeier, J.T., 2001, *An Invariant Approach to the Statistical Analysis of Shapes*. Chapman and Hall/CRC, Boca Raton.

Lewis, P.H., Goodson, K.J., 1990, Images, databases and edge detection for archaeological object drawings.*Computer Applications and Quantitative Methods in Archaeology*, 1990: 149–153.

Leymarie, F.F., 2003, *Three-Dimensional Shape Representation via Shock Flows.* Ph.D. Dissertation. Brown University, Division of Engineering. Providence (RI). (http://www.lems.brown.edu/shape/Presentations/Leymarie02/FolLeymariePhD.pdf). Downloaded on april 2009.

Leyton, M., 1992, *Symmetry, Causality, Mind.* The MIT Press, Cambridge, MA.

Leyton, M., 2005, Shape as memory Storage. In *Ambient intelligence for scientific discovery*. Edited by C. Young. Berlin, Springer-Verlag Lecture Notes in Artificial Intelligence Vol. 3345.

Liming, G., Hongjie, L., Wilcock, J., 1989, The analysis of ancient Chinese pottery and porcelain shapes: A study of classical profiles from the Yangshao culture to the Qing dynasty using computerized profile data reduction, cluster analysis and fuzzy boundary discrimination. *Computer Applications and Quantitative Methods in Archaeology*, 1989: 363–374.

Lohse, E.S., Schou, C., Schlader, R., Sammons, D., 2004, Automated Classification of Stone projectile Points in a Neural Network. In *Enter the Past. The e-way into the four dimensions of culture heritage.* Edited by Magistrat der Stadt Wien-Referat Kulturelles Erbe-Städtarhchäologie Wien. ArcheoPress, Oxford, BAR International Series S1227, pp. 431–437.

Lu, Y. Gardner, H., Jin, H., Liu, N., Hawkins, R., Farrington, I., 2007, Interactive reconstruction of archaeological fragments in a collaborative environment digital image computing techniques and applications. *9th Biennial Conference of the Australian Pattern Recognition Society on Volume*, (3): 23–29.

Lycett, S.J., 2008, Acheulean variation and selection: Does handaxe symmetry fit neutral expectations? *Journal of Archaeological Science*, 35: 2640–2648.

Lycett, S.J., von Cramon-Taubadel, N., Foley, R.A., 2006, A crossbeam co-ordinate caliper for the morphometric analysis of lithic nuclei: A description, test and empirical examples of application. *Journal of Archaeological Science*, 33: 847– 861.

Lyman, R.L., VanPool, T.L., O'Brien, M.J., 2008, Variation in north American dart points and arrow points when one or both are present. *Journal of Archaeological Science*, 35: 2805–2812.

Maaten, L.P.J., Boon, P.J., Lange, A.G., Paijmans, P.P., Postma, E.O., 2006, Computer Vision and Machine Learning for Archaeology. In *Digital Discovery. Exploring New Frontiers in Human Heritage. Fargo CAA2006. Computer Applications and Quantitative Methods in Archaeology*. Edited by J.T. Clark and E.M. Hagemeister, pp. 361–367.

Maaten, L.P.J., Lange, G., Boon, P., 2009, Visualization and Automatic Typology Construction of ceramic profiles. In *Making History Interactive. Proceedings of the 37th Computer Applications in Archaeology Conference*. Edited by B. Frischer and L. Fischer. Williamsburg, VA, March 22–26, 2009. (in press).

Mafart, B., Delingette, H., eds., 2002, Three-Dimensional Imaging in Paleoanthropology and Prehistoric Archaeology, *Acts of the XIVth UISPP Congress,* University of Liège, Belgium, 2–8 September 2001, ArcheoPress, Oxford, BAR International Series, S1049.

Maiza, C., Gaildrat, V., 2005, Automatic Classification of Archaeological Potsherds. In *The 8th International Conference on Computer Graphics and Artificial Intelligence, 3IA′2005,* Limoges, France.

Maiza, C., Gaildrat, V., 2006, SemanticArcheo: A Symbolic approach of Pottery Classification. In T*he 7th International Symposium on Virtual Reality, Archaeology and Cultural heritage. VAST2006.* Edited by M. Ioannidis, D. Arnold, F. Nicolucci and K. Mania. Archeolingua, Budapest.

MacLeod, N., 2002, Geometric morphometrics and geological shape-classification systems. *Earth-Science Reviews*, 59: 27–47.

Mameli, L., Barceló, J.A., Estévez, J., 2002, The Statistics of Archeological Deformation Processes. A zooarchaeological experiment. In *Archaeological Informatics: Pushing the Envelope*. Edited by G. Burenhult. ArcheoPress, Oxford, BAR International Series 1016, pp. 221–230. (http://prehistoria.uab.cat/Barcelo/publication/MameliBarceloEstevez.pdf).

Manferdini, A.M., Remondino, F., Baldissini, S., Gaiani, M., Benedetti, B., 2008, 3D modeling and semantic classification of archaeological finds for management and visualization in 3D archaeological databases. In *Proceedings of 14th International Conference on Virtual Systems and MultiMedia (VSMM 2008)*, pp. 221–228, Cyprus, Limassol.

Mara, H., Sablatnig, R., 2005a, A Comparison of Manual, Semiautomatic and Automatic Profile Generation for Archaeological Fragments. In *Proceedings of the 10th Computer Vision Winter Workshop*. Edited by A. Hanbury, H. Bischof. Zell an der Pram, Austria, pp. 123–134.

Mara, H., Sablatnig, R., 2005b, 3D-vision applied in archaeology. Forum Archaeologiae – Zeitschrift für klassische Archäologie, 34(3).

Mara, H., Sablatnig, R., 2006, Orientation of Fragments of Rotationally Symmetrical 3D-Shapes for Archaeological Documentation. In *Proceedings of 3rd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT).* Edited by M. Pollefeys, K. Daniilidis. Chapel Hill, June 14–16, pp. 1064–1071.

Mara, H., Sablatnig, R., 2008, Evaluation of 3D shapes of ceramics for the determination of manufacturing techniques. In layers of perception. *Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology*, Berlin, April 2007. Edited by A. Posluschny, K. Lambers and I. Herzog. Berlin, Rudof von Habelt Verlag (Kolloquien zur Vor- und Frühgeshichte, Band 10.

Mara, H., Sablatnig, R., Karasik, A., Smilansky, U., 2004, The Uniformity of Wheel Produced Pottery Deduced from 3D Image Processing and Scanning. In *Digital Imaging in Media and Education, Proceedings of the 28th Workshop of the Austrian Association for Pattern Recognition* (OAGM/AAPR). Edited by W. Burger, J. Scharinger. Schriftenreihe der OCG, Vol. 179, pp. 197–204.

Marr, D.H., 1982, *Vision, A Computacional Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco.

Martin, D.H., Fowlkes, C., Malik, J., 2004, Learning to 1detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 26(5): 530–549.

Martín, D., Melero, F.J., Cano, P., Torres, J.C., 2009, Feature preserving simplification of Point Clouds from large Range laser Scanners. In *Making History Interactive. Proceedings of the 37th Computer Applications in Archaeology Conference*. Edited by B. Frischer and L. Fischer. Williamsburg, VA, March 22–26, 2009. (in press).

Marr, D., Hildreth, E., 1980, Theory of edge detection. *Proceedings of the Royal Society of London B*, 207: 187–217.

McClamrock, R., 1993, Functional analysis and Etiology. *Erkentniss*, 38: 249–260.

McGrew, W.C., 1993, The Intelligent Use of Tools: Twenty Propositions. In *Tools, Language, and Cognition in Human Evolution*. Edited by K.R. Gibson and T. Ingold. Cambridge University Press, Cambridge, pp. 151–170.

Melero, F.J., Torres, J.C., Leon, A., 2003, On the Interactive 3d Reconstruction of Iberian Vessels. In *4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, VAST'03*. Edited by F. Niccolucci D. Arnold and A. Chalmers.

Meltzer, D.J., Cooper, J.R., 2006, On morphometric differentiation of clovis and non-clovis blades. *Current Research in the Pleistocene*, 23: 143–145.

Millikan, R.G., 1999, Wings, spoons, pills and quills: A pluralist theory of function. *Journal of Philosphy*, 96: 192–206.

Mom, V., 2005, SECANTO –The Section Analysis Tool. In *The World is in your Eyes. Proceedings of the XXXIII Computer Applications and Quantitative methods in Archeology Conference* (March 2005). Edited by A. Figueiredo and G. Velho. Tomar, Portugal, pp. 95–102. Associaçâao para o Desenvolvimiento das Aplicaçôes Informáticas e Novas Tecnologias em Arqueologia.

Mom, V., 2006, Where Did I See You Before. Holistic Method to Compare and Find Archaeological Artifacts. In *Advances in Data Analysis. Proceedings of the 30th Annual Conference of the Gesellschaft für Klassifikation* e.V. Edited by R. Decker and H.-J. Lenz. Springer, Freie Universität Berlin, March 8–10.

Mom, V., Paijmans, H., 2008, SECANTO: A Retrieval System and Classification Tool for Simple Artefacts. In *Layers of Perception. Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA).* Berlin, Germany, April 2–6, 2007. Edited by A. Posluschny, K. Lambers and I. Herzog, Kolloquien zur Vor- und Frühgeschichte, Vol. 10. Bonn Dr. Rudolf Habelt GmbH.

Morris, G., Scarre, C.J., 1981, Computerized analysis of the shapes of prehistoric stone tools from west-central France. *Computer Applications in Archaeology*, 1981: 83–94.

Movchan, A.B., Movchan, N.V., 1998, *Mathematical Modeling of Solids with Nonregular Boundaries*. CRC Press, London.

Nagel, E., 1961, *The Structure of Science*. Harcourt, Brace and World Inc., New York and Burlingame.

Nautiyal, V., Kaushik, V.D., Pathak, V.K., Dhande, S.G., Nautiyal, S., Naithani, M., Juyal, S., Gupta, R.K., Vasisth, A.K., Verma K.K., Singh, A., 2006, Geometric Modeling of Indian Archaeological Pottery: A Preliminary Study. In*Digital Archaoelogy. Exploring new frontiers in Human Heritage.* Edited by J.T. Clark and E.H. Hagemeister. Archeolingua press, Budapest.

Neander, K., 1991, The teleological notion of function. *Australian Journal of Philosophy*, 69: 454–68.

Nelson, B., 1985, Reconstructing Ceramic Vessels and Their Systemic Contexts. In *Decoding Prehistoric Ceramics*. Edited by B. Nelson. Southern Illinois University Press, Carbondale, pp. 310–329.

Nelson, R.C., Selinger, A., 1998, Large-scale tests of a keyed, appearance-based 3-D object recognition system. *Vision research*, 38(15–16): 2469–2488.

Nowell, A., Park, K., Metaxas, D., Park, J., 2003, Deformation Modeling: A Methodology for the Analysis of Handaxe Morphology and Variability. In *Multiple Approaches to the Study of Bifacial Technologies*. Edited by M. Soressi and H.L. Dibble. University of Pennsylvania Museum of Archaeology and Anthropology, Philadelphia, pp. 193–208.

O'Brien, M.J., Lyman, R.L., 2002, *Seriation, Stratigraphy and Index Fossils*. Kluwer, New York.

O'Gorman, L., Sammon, M.J., Seul, M., 2008, *Practical Algorithms for Image Analysis.* Cambridge University Press, Cambridge.

Orton, C., Tyers, P., Vine, A., 1993, *Pottery in Archaeology.* Cambridge University Press, Cambridge.

Otárola-Castillo, E., Adams, D., Coinman, N.R., Collyer, M.C., 2008, Differences between Point Morphology Methods: Traditional Caliper Measurements versus Geometric Morphometrics. In *Advances in the Study of Lithic Morphology Symposium*. 73rd Annual Society for American Archaeology Meeting, Vancouver, British Columbia, Canada, March 27. Edited by E. Otárola-Castillo, B.J. Schoville and A.R. Boehm. ArcheoPress, Oxford (in press).

Ozmen, C., Balcisoy, S., 2006, 3D Spatial Measurement Tools for Digitized Artifacts. In *Digital Archaoelogy. Exploring new frontiers in Human Heritage.* Edited by J.T. Clark and E.H. Hagemeister. Archeolingua press, Budapest.

Palmer, S., 1999, *Vision Science. Photons to Phenomelogy*. The MIT Press, Cambridge, MA.

Pande, C.S., Richards, L.R., Smith, S., 1987, Fractal characteristics of fractured surfaces. *Journal of Materials Science Letters*, 6(3).

Papaioannou, G., Karabassi, E.A, 2003, On the automatic assemblage of arbitrary broken solid artefacts. *Image and Vision Computing*, 21: 401–412.

Papaioannou, G., Karabassi, E.A., Theoharis, T., 2002, Reconstruction of three-dimensional objects through matching of their parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1): 114–124.

Pasko, G., Pasko, A., 2006, Function-based Shape modelling fort Cultural heritage Applications. In *The 7th International Symposium on Virtual Reality, Archaeology and Cultural heritage. VAST2006.* Edited by M. Ioannidis, D. Arnold, F. Nicolucci and K. Mania. Archeolingua, Budapest.

Pasko, A., Adzhiev, A., Sourin, V., Savchenko, V., 1995, Function representation in geometric modeling: Concepts, implementation and applications. *The Visual Computer*, 11(8): 429–446.

Pechuk, M., Soldea, O., Rivlin, E., 2005, Function-Based Classification from 3D Data via Generic and Symbolic Models. In *Twentieth National Conference on Artificial Intelligence (AAAI-05),* Pittsburgh, PA (http://www.cs.technion.ac.il/˜mpechuk/publications/oclsAAAI05.pdf).

Peng, L.W., Shamsuddin, S.M., 2004, Modeling II: 3D Object Reconstruction and Representation using Neural Networks. In *Proceedings of the 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast GRAPHITE '04.* Published by the Academy of Computing Machinery Press.

Petersen, C., Schlader, R., Chapman, R.E., Deck, L.T., Clement, N., Heydt, R., 2006, Three-Dimensional Scanning of Archeological Objects for Research, Outreach, and Specimen Archiving: Potentials and Responsibilities. In *Digital Archaoelogy. Exploring new frontiers in Human Heritage.* Edited by J.T. Clark and E.H. Hagemeister. Archeolingua press, Budapest.

Peterson, J.W.M., 1992, Fourier Analysis of Field Boundaries In *CAA91: Computer Applications and Quantitative Methods in Archaeology 1991* BAR International Series S577. Edited by G. Lock and J. Moffett. Tempus Reparatum, Oxford, pp. 149–156.

Pijoan, J., 2007, *Quantificació de traces d'us en instruments lítics mitjançant imatges digitalitzades: Resultats d'experiments amb Xarxes Neurals i estadística.* Ph.D. Dissertation. Universitat Autonoma de Barcelona, Spain.

Piperakis, E., Kumazawa, I., 2001, Affine Transformations of 3D Objects Represented with Neural Networks. In *3-D Digital Imaging and Modeling, Proceedings*, pp. 213–223.

Pires, H., Ortiz, P., Marques, P., Sanchez, H., 2007, Close-range Laser Scanning Applied to Archaeological Artifacts Documentation. Virtual Reconstruction of an XVIth Century Ceramic Pot. In *The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2006).* Edited by M. Ioannides, D. Arnold, F. Niccolucci and K. Mania. ArcheoLingua, Budapest.

Pobelome, J. et al. 1997, Manual Drawing Versus Automated Recording of Ceramics. In: *Sagalasos IV. Acta Archaeologica Lovaniensia Monographiae 9*. Edited by M. Walkens. Leuven, pp. 533–538.

Ponce, J., Hebert, M., Schmid, C., Zisserman, A., 2007, *Toward Category-Level Object Recognition,* Springer-Verlag Lecture Notes in Computer Science, Vol. 4170.

Porter, D., Werner, P., Utcke, S., 2005, Ancient Ceramics: Computer aided Classification. *Technical Report M-338*. University of Hamburg, Computer Science Department, KOGS – Cognitive Systems Group. (http://www.informatik.uni-hamburg.de/bib/medoc/M-338.pdf). Downloaded on April 2009.

Puente, C.E., Castillo, P.A., 1996, On the fractal structure of networks and dividers within a watershed. *Journal of Hydrology*, 187: 173–181.

Razdan, A., Liu, D., Bae, M., Zhu, M., Farin, G., Simon, A., Henderson, M., 2001, Using Geometric Modeling for Archiving and Searching 3D Archaeological Vessels. *CISST 2001* June 25–28, 2001, Las Vegas.

Razdan, A., Liu, D., Bae, M., Zhu, M., Simon, A., Farin, G., Henderson, M., 2004, Shape Modeling for 3D Archaeological Vessels. In *Geometric Modeling: Techniques, Applications, Systems and Tools*. Edited by M. Sarfraz. Kluwer Academic Publishers, Norwell, MA, pp. 362–374.

Read, D., 2007, *Artifact Classification. A Conceptual and Methodological Approach.* Left Coast Press, Walnut Creek, CA.

Reuter, P., Riviere, G., Couture, N., Sorraing, N., Espinasse, L., Vergnieux, R., 2007, ArcheoTUI – A Tangible User Interface for the Virtual. In *The 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage. VAST (2007)* Edited by D. Arnold, F. Niccolucci and A. Chalmers. ArcheoLingua, Budapest.

Richards, J.D., 1987, *The Significance of Form and Decoration of Anglo-Saxon Cremation Urns.* ArcheoPress, Oxford (British Archaeological Reports, British Series, 166).

Riel-Salvatore, J., Bae, M., McCartney, P., Razdan, A., 2002, Paleolithic archaeology and 3D visualisation technology: Recent developments. *Antiquity*, 76: 929–930.

Rovner, I., 1993, Complex Measurements Made Easy: Morphometric Analysis of Artefacts Using Expert Vision Systems. In *Computer Applications in Archaeology 1993.* Edited by J. Wilcock and K. Lockyear. ArcheoPress, Oxford (British Archaeological Reports).

Rovner, I., 2006, Computer-assisted Morphometry of Digital Images: Beyond Typology in the Morphological Analysis of the Broad Spectrum of Archaeomaterials. In *Digital Archaoelogy. Exploring New Frontiers in Human Heritage*. Edited by J.T. Clark and E.H. Hagemeister. Archeolingua press, Budapest.

Rovner, I., Gyulai, F., 2007, Computer-assisted morphometry: A new method for assessing and distinguishing morphological variation in wild and domestic seed populations. *Economic Botany*, 61(2): 154–172.

Rowe, J., Razdan, A., 2003, A Prototype Digital Library For 3D Collections: Tools To Capture, Model, Analyze, and Query Complex 3D Data. *Museums and the Web 2003 Conference,* March 19–22, Charlotte, North Carolina.

Rushmeier, H., Xu, C., Wang, B., Rushmeier, R., Dorsey, J., 2007, Shape Capture Assisted by Traditional Tools. In *8th International Symposium on Virtual reality, Archaeology and Intelligent Cultural Heritage (5th Eurographics Workshop on Graphics and Cultural heritage),* Brighton, November 26–30, 2007. Published by Eurographics Association, TU Darmstadt & Fraunhofer IGD, pp. 1–8.

Russ, J.C., 1990, *Computer-Assisted Microscopy: The Measurement and Analysis of Images*. Plenum Press. New York.

Russ, J.C., 2006, *The Image Processing Handbook*, CRC Press, London.

Russ, J.C., Rovner, I., 1989, Expert vision systems in archaeometry: Rapid morphometric analysis of chaotic form, shape and structure. *Materials research society bulletin: Microscopy for the archaeologist*, 14: 3140–144.

Sagiroglu, M., Erçil A., 2005, A Texture Based Approach to Reconstruction of Archaeological Finds. In *Proceedings of VAST 2005*, pp. 137–142.

Saragusti, I., Karasik, A., Sharon, I., Smilansky, U., 2005, Quantitative analysis of shape attributes based on contours and section profiles in artifact analysis.*Journal of Archaeological Science*, 32(6): 841–853.

Sarfraz, M., 2007, *Interactive Curve Modeling: With Applications to Computer Graphics, Vision and Image Processing*. Springer, New York.

Schneidermann, H., Kanade, T., 2004, Object detection using the statistics of parts. *International Journal of Computer Vision*, 56(3): 151–177.

Schurmans, U., Razdan, A., Simon, A., Marzke, M., McCartney, P., Van Alfen, D., Jones, G., Zhu, M., Liu, D., Bae, M., Rowe, J., Farin, G., Collins, D., 2002, Advances in Geometric Modeling and Feature Extraction on Pots, Rocks and Bones for Representation and Query via the Internet. In *Archaeological Informatics: Pushing the Envelope CAA 2001. Computer Application and Quantitative Methods in Archaeology. Proceedings of the 29th Conference*, Gotland, April 2001. Edited by G. Burenhult and J. Arvidsson. Archaeopress, Oxford, BAR International Series 1016, pp. 191–202.

Senior, L.M., Birnie, D.P., 1995, Accurately estimating vessel volume from profile illustrations, *American Antiquity*, 60(2): 319–334.

Shelley, C.P., 1996, Visual abductive reasoning in archaeology. *Philosophy of Science*, 63: 278–301.

Simon, A., Van Alfen, D., Razdan, A., Farin, G., Bae, M., Rowe, J., 2002, 3D Modeling for Analysis and Archiving of Ceramic Vessel Morphology: A Case Study from the

American Southwest. In *Proceedings of the 33rd International Symposium on Archaeometry. Geoarchaeological and Bioarchaeological Studies,* Vrije Universiteit, Amsterdam, 2002.

Slice, D.E., 2007, Geometric morphometrics. *Annual Review of Anthropology*, Vol. 36.

Small, C.G., 1996, *The Statistical Theory of Shape*. Springer, New York.

Smith, M.F., 1983, *The Study of Ceramic Function from Artifact Size and Shape*. PhD. Disertation. University of Oregon.

Smith, M.F., 1985, Toward an Economic Interpretation of Ceram- ics: Relating Vessel Size and Shape to Use. In *De- coding Prehistoric Ceramics*. Edited by B. Nelson. Southern Illinois University Press, Carbondale, pp. 254–309.

Sonka, M., Hlavac, V., Boyle, R., 1984, *Image Processing, Analysis and Machine Vision*. Chapman and Hall, London.

St. Amant, R., 2002, A Preliminary Discussion of Tools and Tool Use'. *Technical Report TR-2002-06,* North Carolina State University.

Steine, M., 2005, *Approaches to Archaeological illustration. A Handbook.* Council for British Archaeology. Practical Handbook No. 18.

Trinkl, E., 2005, 3d-vision applied in archaeology. *Forum Archaeologiae Zeitschrift für klassische Archäologie*, 34(III): 2005.

Tsioukas, V., Patias, P., Jacobs, P.F., 2004, A novel system for the 3D reconstruction of small archaeological objects. *International Archives Of Photogrammetry Remote Sensing And Spatial Information Sciences*, 35(5): 815–818.

Tsirliganis, N., Pavlidis, G., Koutsoudis, A., Papadopoulou, D., Tsompanopoulos, A., Stavroglou, K., Loukou, Z., Chamzas, C., 2002, Archiving 3D Cultural Objects with Surface Point-Wise Database Information, In *First International Symposium on 3D Data Processing Visualization and Transmission*, June 19–21, 2002, Padova, Italy.

Tuceryan, M., Jain, A.K., 1993, Texture Analysis. In *Handbook of Pattern Recognition & Computer Vision*. Edited by C.H. Chen, L.F. Pau and P.S.P. Wang. World Scientific, Singapore, pp. 235–276.

Van der Leeuw, S.E., 2000, Making Tools from Stone and Clay. In *Australian archaeologist: Collected papers in honour of Jim Allen*. Edited by P. Anderson and T. Murray. Australian National University Press, Canberra.

Vermaas, P.E., Houkes, W., 2003, Ascribing functions to technical artefacts: A challenge to etiological accounts of functions. *British Journal for the Philosophy of Science*, 54: 261–289.

Vilbrandt, C.W., Pasko, G.I., Pasko, A.A., Fayolle, P.A., Vilbrandt, T.I., Goodwin, J.R., Goodwin, J.M., Kunii, T.L., 2004, Cultural heritage preservation using constructive shape modeling. *Computer Graphics Forum*, 23(1): 25–41.

Wang, H., Li, H., Ye, X., Gu, W., 2000, Training a neural network for moment based image edge detection. *Journal of Zhejiang University SCIENCE*, 1(4): 398–401.

Whallon, R., 1982, Variables And Dimensions: The Critical Step In Quantitative Typology. In *Essays In Archaeological Typology*. Edited by R. Whallon and J.A. Brown. Center for American archaeology press, Evanston, IL, pp. 127–161.

Wilcock, J.D., Shennan, S.J., 1975, Computer Analysis of Pottery Shapes. In *Computer Applications in Archaeology 1975*. Edited by S. Laflin. University of Birmingham, England, pp. 98–106.

Willemin, J.H., 2000, Hack′s Law: Sinuosity, convexity, elongation. *Water Resource Research*, 36: 3365–3374.

Willis, A.R., Cooper, D.B., 2004, Bayesian Assembly of 3d Axially Symmetric Shapes from Fragments. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 82–89.

Willis, A., Orriols, X., Cooper, D.B., 2003, Accurately Estimating Sherd 3D Surface Geometry with Application to pot Reconstruction. In *IEEE Workshop on Applications of Computer Vision in Archaeology*.

Winkelbach, S., Wahl, F.M., 2008, Pairwise matching of 3D fragments using cluster trees. *International Journal of Computer Vision*, 78(1): 1–13.

Wright, L., 1973, Functions. *Philosphical Review*, 82: 139–68.

Wynn, T., Tierson, F., 1990, Regional comparison of the shapes of later Acheulan handaxes. *American Anthorpologist*, 92: 73–84.

Yao, F., Shao, G., 2003, Detection of 3D Symmetry Axis from Fragments of a Broken Pottery Bowl. Acoustics, Speech, and Signal Processing. *Proceedings (ICASSP '03)*, 3: 505–508.

Zaharieva, M., Kampel, M., Vondrovec, K., 2008, From Manual to Automated Optical Recognition of Ancient Coins. In *VSMM 2007*. Edited by T.G. Wyeld, S. Kenderdine and M. Docherty. LNCS 4820, 2008. Springer, Berlin, pp. 88–99.

Zhang, W.Y., Tor, S.B., Britton, G.A., 2002, Automated functional design of engineering systems. *Journal of Intelligent Manufacturing*, 13(2): 119–133.

Zhou, M., Wu, Z., Shui, W., 2009, Computer Assisted Recovery Technology of Broken Rigid Objects and Its Applications in Terra Cotta Warriors and Horses. In *Making History Interactive*. *Proceedings of the 37th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA)*. Edited by B. Frischer and L. Fisher. Williamsburg, VA, March 22–26, 2009. (in press).

Zhu, L., Zhou, Z., Zhang, J., Hu, D., 2006, A Partial Curve Matching Method for Automatic Reassembly of 2D Fragments. In Intelligent Computing in Signal Processing and Pattern Recognition. Edited by D.-S. Huang, K. Li, and G.W. Irwin. Springer, LNCIS 345, pp. 645–650.

Chapter 6
Evolution of the Upper Cretaceous Oysters: Traditional Morphometrics Approach

Ahmed A. Abdelhady and Ashraf M.T. Elewa

Idea and Aims

The faunal association, which is revealed by newly discovered, well preserved, and therefore highly diverse assemblages, enabled us to recognize some phylogenetic trends in the Upper Cretaceous oysters. In the present study, the oyster faunal collection from El-Sheik Fadl-Ras Gharib stretch, Eastern Desert, Egypt, identifies some of the processes by which distinct species may have transformed into another as a result of the change in the environmental conditions. The applied qualitative and quantitative techniques exhibited considerable stability in assigning the relationships among taxa.

Introduction

The origin of oysters is a much debated palaeontological issue. The knowledge of their total ranges and of evolutionary relationships still fragmentary; despite the Late Cretaceous oyster data have became available. The origin and taxonomic status of oysters (Ostreidae Rafinesque 1815) and oyster-like cementing bivalves, namely Plicatulidae and Dimyidae, have traditionally been controversial. Most hypotheses of relationship between the three families (and corresponding superfamilies) have relied on soft and hard part synapomorphies shared by recent taxa. Waller (1978) placed Dimyoidea, Plicatuloidea, and Ostreoidea in his suborder Ostreina (=Ostreoida). Alternatively, Yonge (1978, 1980) believed that the Dimyidae was not related to the Ostreidae or Pectinidae, but that they were likely ancestors of the Plicatulidae. Hautmann (2001) rejected Waller's hypothesis of relationships for the following three reasons:

A.A. Abdelhady (✉)
Geology Department, Minia University, Minia, Egypt
e-mail: alhady2003@yahoo.com

(1) Ostreidae attach themselves to the substrate by their left valve, while Plicatulidae and Dimyidae are dextrally attached,
(2) The inner shell layer of early oysters was probably nacreous, but crossed lamellar in Dimyidae and Plicatulidae, and
(3) Secondary hinge teeth and secondary hinge ligament are absent in Ostreidae.

The opening of seaways by sea-floor spreading in the Atlantic led to a major global transgression in the Mid-Cretaceous. This is supposed to be the main reason for the generally observed, increasingly cosmopolitan character of bivalve faunas at the very end of the Early Cretaceous. Rising sea level eliminated geographical barriers, although it has been shown that sea-level rise probably also led to provincialism at that time. Taking into account the observations of earlier studies on oysters concerning constraints on the use of fossil data in palaeobiogeographical analyses, the invertebrate benthic fauna, occasionally oysters, are particularly useful because:

1. They seem to constitute a monophyletic group, and the taxa can be identified with high degree of precision; they are abundant in samples from facies favorable for their preservation;
2. They have a relatively complete fossil record and their environmental requirements are well known;
3. The stratigraphical framework of the sections containing sediments rich in oysters is well known, making it possible to construct a sequence of palaeobiogeographical charts for the Cretaceous;
4. The ages of the taxa and the available palaeogeographical reconstructions are closely linked.

Consequently, phylogeny is essential for reconstructing the taxonomic, environmental and biogeographical patterns of this group, and to understanding the causal processes responsible for them. Several authors dealt with the Upper Cretaceous oysters and their taxonomy, ecology and biogeography (i.e. Mancini 1978; Aqrabawi 1993; Malchus 1996; Seeling and Bengston 1999; Dhondt 1999, Dhondt and Jaillard 2005). Even though, the evolution of the Upper Cretaceous Exogyrinae oysters received little attention.

In the current study, we have two major goals:

1. To examine the morphological characters of the oyster group in order to evaluate the old phylogenetic schemes,
2. To assess the migration routes and other biogeographic relations between the different biotic provinces.

Previous Phylogenetic Schemes

The phylogeny of Upper Cretaceous oysters is firstly examined by Böse (1919). He thought that populations of *Exogyra plexa* and *IIymatogyra arietina*, which are occurring at different stratigraphical levels, were sufficiently distinguishable; leading to the result of regarding some of them as new species (Fig. 6.1). Recently,

Fig. 6.1 Modified phylogeny tree of Texas exogyrid oysters proposed by Böse (1919), Genera are updated according to Moore (1969)

Maasterichtian	E. costata		
Campanian	E. panderosa		
Santonian	E. aff panderosa		E. laevluscala
Coniacian			
Turonian			E. columbella
Cenomanian	E. clarki / E. drakel / E. whltneyi / E. cartiedgle	E. fenox	I. arietina / I. sp.n.aff.arietina
Albian	Amphidonte walkeri / C. texanum / Exogyra paupercula	E. plexa / C. weatherfondensis	E. aff plexa / E. plexa sp.aff plexa / E. plexa n.sp.aff plexa
Aptian	Ceratostreon hill		

Malchus (1990) proposed a phylogeny, based on morphological and microstructure, showing that *E. plexa* lineage branches from the *Ceratostreon* lineage, and then giving rise to a major radiation of exogyrid oysters. Moreover Malchus noted that there are intermediate forms occuring between *E. plexa* and *I. arietina* from the strata of North America and North Africa (Fig. 6.2). Yurke and Charles (2006) made reference collections of populations of *E. plexa* and *I. arietina* from different stratigraphical levels similar to those examined by Böse (1919). They extended the genus *Ilymatogyra* to include *E. plexa* and *E. cartledgei* (Fig. 6.3), but they could not identify the ancestor of *Ceratostreon texanum* or *Exogyra plexa.*

From the above mentioned studies, we conclude that:

(1) There were misidentifications for most of the Exogyrinae oysters,
(2) Some taxa exhibit dimorphism or even polymorphism,
(3) The great abundances occur within the Cenomanian genera, while Turonian/Coniacian genera have the minimum diversity indexes.

Material and Methods

Six Stratigraphical sections are chosen from the study area according to their richness on macrofauna especially oysters. These sections extend from the SE to NW along the El-Sheik Fadl-Ras Gharib stretch, Eastern Desert, Egypt. They were

Fig. 6.2 A phylogeny of exogyrid oysters; proposed by Malchus (1990). The range of each oyster species is indicated by a *thick vertical black bar*
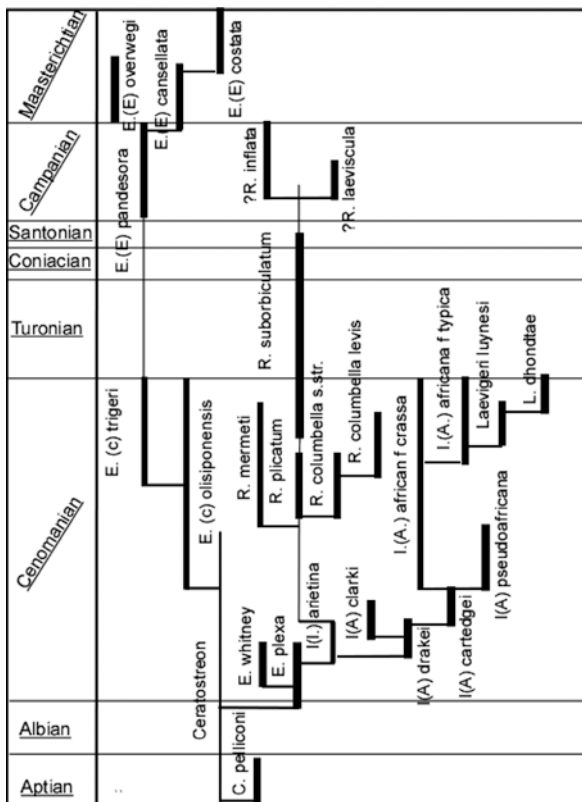
measured in the field as a bed by bed using handy lens, and their lithologies, sedimentary patterns and fossil occurrences were described (Fig. 6.4). Lithological and biostratigraphical data from these sections are herein presented. A total of 21 species were identified in the present study. The typical representatives of the whole assemblage are deposited at the Geological Museum of the Geology Department, Faculty of Science, Minia University, Egypt (ABO1: ABO21).

The study is based on a combination of fossil species identified from the field collections (Fig. 6.4), and from literatures. Fauna prepared, photographed, and a special emphasis is given to the morphological data for analytical methods concerning evolution.

Previous examinations of shell morphologies of taxa revealed unsuspected morphological differences, among *Exogyra*, *Ceratostreon*, *Rhynchestreon* and *IIymatogyra*; as all taxa have a wide range of shell pattern variation. Specimens were randomly selected from each location, and the shells photographed with a Sony DH9 digital camera (12 Mega pixel, 15X) or scanned with an HP 4870 Pro flatbed scanner. Only left valves were used in the analyses because of the inequivalve hinge morphology between left and right valves.

Fig. 6.3 Recent phylogeny of exogyrid oysters; proposed by Yurke and Charles (2006)
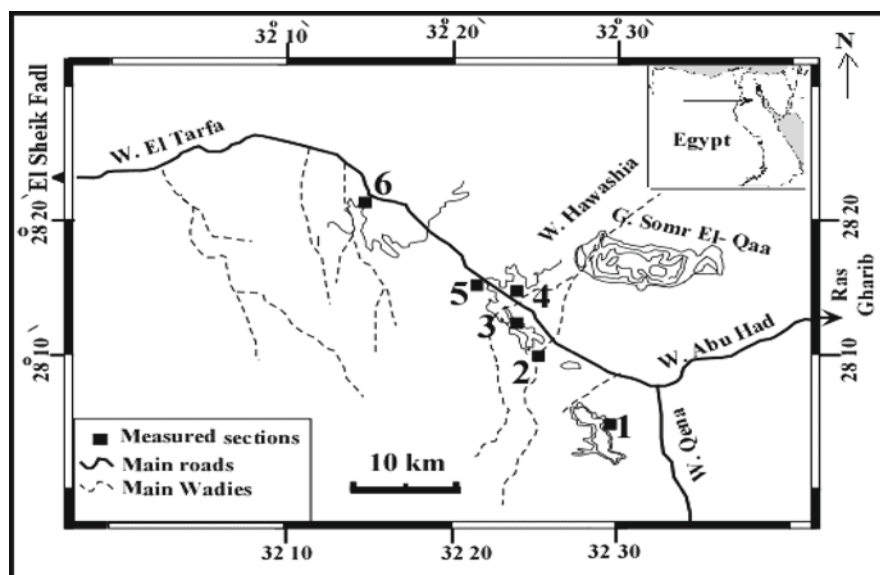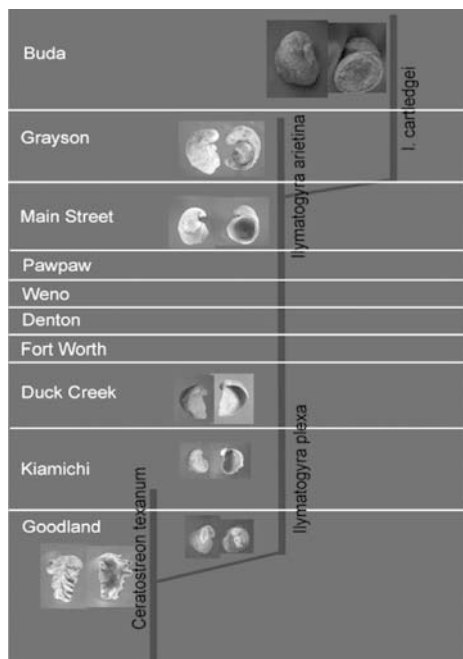
Fig. 6.4 Location map of the studied sections

Fig. 6.5 Internal view of left valve, showing measurements carried on the oyster shells
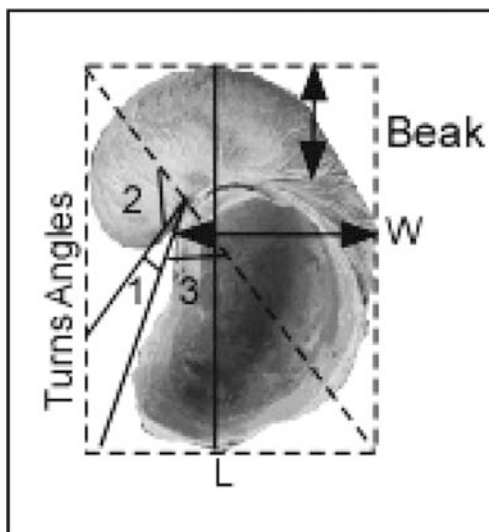
In order to explore the morphological affinities of Exogyrinae, we examined the distribution of qualitative morphological characters for 10 species from 4 genera of Exogyrinae oysters. Following preliminary analyses, the data comprise 9 qualitative characters (4 dimensions, 3 angles and 2 ratios) (Fig. 6.5). These variables are chosen in order to minimize the morphological proxies, for individual section a number of shell photographed, scanned then coordinates and angles were digitally captured from specimen images using the software TPSdig of Rohlf (1998). The obtained measures were analyzed using the PAST program, version 1.72, where we applied all the required quantitative techniques (for details see Oyvind et al. 2001).

Qualitative Observations

Important Criteria

The short comings of deriving phylogeny from the first and last appearances of taxa and the incompleteness of the fossil record have been widely discussed (Smith and Bengtson 1991). Recognition of attachment orientation is a very important criterion derived from the fact that oysters and other cementers (Prospondylidae, Plicatulidae, Dimyidae) usually have a slightly opisthogyrate umbo (Fig. 6.6), and the valve, as a whole, twists towards the posterior side, unless growth is hindered by irregularities of the substrate (Márquez-Aliaga et al. 2005). This character was maximally developed in Exogyrinae oysters and was not so evident in other cementers closely connected to pectinoids, such as Spondylidae. The stratigraphical occurrence also represents an important criterion, the vertical change in ecological parameters suggests the occurrence of related morphotypes, species and genera.

Exogyriine	Pycnodontiide	Gyrostriidae	Osteriidae
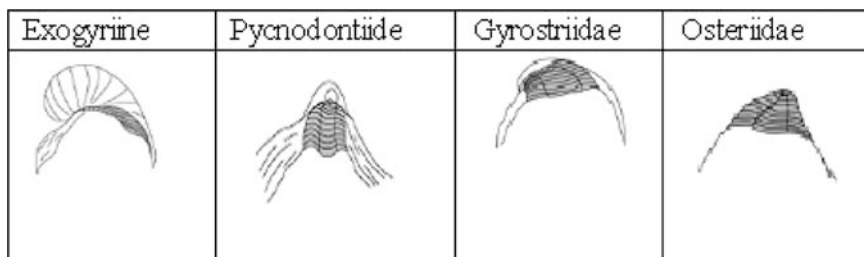
Fig. 6.6 Schematic key, based on umbonal area, for oysters differentiation

Distribution of specific taxa is influenced by environmental parameters. Some taxa have the ability to tolerate these parameters by giving rise to a morphotype with modification of some of its old morphological characters. It is important to note that the whole morphology of the morphotype is still governed by the close relation to the parent "original" type; these taxa can only be mentioned in evolutionary trends.

A final criterion, which can be used in combination with the above, is the valve convexity. In Pseudomonotidae, Prospondylidae, the majority of Ostreoidea, and many Plicatulidae and Dimyidae (as well as in other cementers) the right valve is less convex than the left one, Abdelhady (2007). In many oysters, both valves have similar convexities. Several hundreds of specimens have been examined from the late Albian to the early Turonian. These specimens were collected in the same states. Shell ornamentation patterns have been used as diagnostic characters for these species; the taxonomic status of which has not yet been revised following Moore (1969). Younger identified genera are also taken into account. By using the criterion of attachment orientation, stratigraphical occurrence, valve turns, valve convexity and beak size in addition to beak twisting, we herein suggest new evolutionary trends.

Body Size Evolution

Body size in Exogyrinae oysters has greatly increased (Fig. 6.8). This trend is evident both in raw data and when populations are binned into million-year intervals. The oldest population in the analysis is also the smallest; while later populations are nearly 50% larger in length. Most of the increase in size is over 8 million years. The temporal trend in body size mirrors the trajectory of bottom-water temperature, which heated most rapidly in the years of this interval (Fig. 6.7). Bottom-water temperature yields a highly positive relationship. Bottom-water temperatures that raised the body size in these oysters have increased. In agreement with the second prediction of the Cope–Bergmann hypothesis, an inverse correlation between temperature and body size is evident.

Intriguingly, over global heating, the average size of late Cenomanian *Exogyra* (*C.*) *olisiponensis*, *Gyrostrea delettrei* and *IIymatogyra areitina* are the larger within Exogyrinae oysters. Dimensions indicate that the individuals of *I.* (*A.*) *Africana*
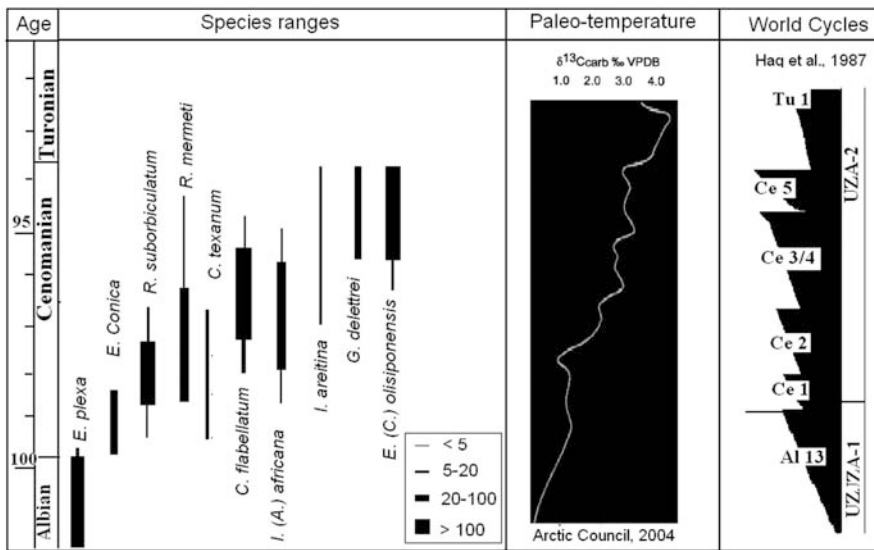
Fig. 6.7 Species ranges of Exogyrinae oysters versus age. The figure shows correlations with the paleo-temperatures of the Arctic Council (2004), and the world cycles of Haq et al. (1987)
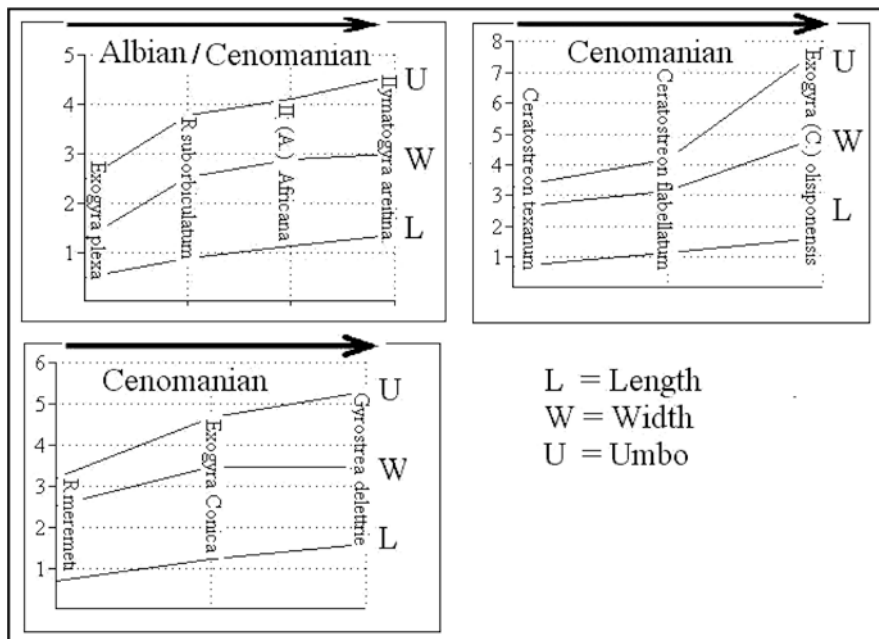
Fig. 6.8 The increase in body size of Exogyrinae oysters with time

exhibit intermediate size proportions when compared to the previously three largest species described. In general term, large body size reflects higher temperate latitudes with higher ocean productivity; herein is represented by the course carbonate beds with minor shales and mud intercalations of the Abu Qada Formation.

The Studied acme Zones

The acme zone takes its name from the taxon that has the greater abundance within its association. The late Albian/Cenomanian oysters of the studied sections led to divide the interval into the following acme Zones:

1. *Exogyra plexa-Exogyra ribose* Acme Zone (late Albian),
2. *Rhyncostreon suborbiculatum* Acme Zone (early Cenomanian),
3. *Ceratostreon flabellatum* Acme Zone (middle Cenomanin), and
4. *Exogyra (Costagyra) olisiponensis* Acme Zone (late Cenomanian).

Stage Boundaries

Albian/Cenomanian

The A/C boundary is defined herein at the LADs of *Exogyra plexa*, and is marked by the FADs of *E. plexa*. No marked depositional hiatuses or sharp changes are reported at this boundary, so it is believed that continuous sedimentation took place within the interval of latemost Albian to Cenomanian, with sudden increase in species diversity and richness of Exogyrinae oysters.

Cenomanian/Turonian

The C/T boundary of the present study is determined by the LAD of *Exogyra* (*Costagyra*) *olisiponensis* and/or *IIymatogyra areitina* (Fig. 6.7), where the oysters go vanish. The latest Cenomanian eustatic transgression thought to have been the most intense Phanerozoic flooding event. Maximum sea level during the C/T interval was probably 255 m higher (Haq et al. 1987), and shelf areas were twice as large as those of today. The levels of the atmospheric CO_2 were at least four times when compared with the present levels. The C/T boundary of the studied sections, therefore, represents a great mass extinction.

Morphometric Analyses

Dimorphism

Dimorphism has been reported to occur in some species of Exogyrinae; therefore we used the principal component analysis to examined two Exogyrinae species that are close to each other in their morphological characters:
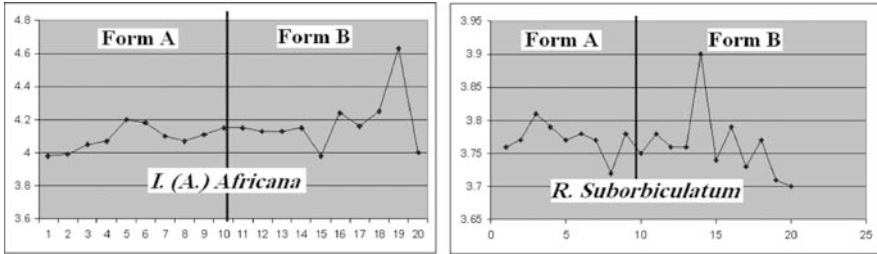
Fig. 6.9 Shell diameters of *I.* (*A.*) *Africana* and *R. suborbiculatum*
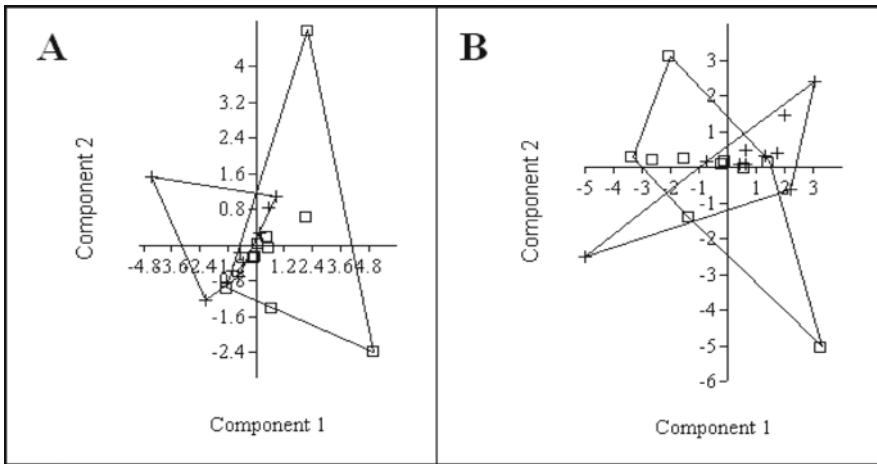
Fig. 6.10 Principle component scatter plot, based on correlation matrix with 95% confidence, for (**a**) *I.* (*A.*) *Africana*; percentage of variation in PC1 = 43.829, and (**b**) *R. suborbiculatum*, percentage of variation in PC1 = 57.82

1. *Ilymatogyra* (*Afrogyra*) *Africana*, and
2. *Rhynchostreon suborbiculatum* (Fig. 6.9)

In Fig. 6.10, the PC scatter plot represents slight variation in morphologies of the two species, but without ability to isolate sub-specific characters. Figure 6.13a clearly shows the dimorphism of *I.* (*A.*) *Africana*; as it is believed by Malchus (1990), who subdivided *I.* (*A.*) *Africana* into two morphs, and suggested that such variation is related to the type of substratum.

For *R. suborbiculatum*, the two morphs are evident. These morphs may indicate two facts; the first is of Moore (1969), which indicates ontogenetic stage variation; while the other is dimorphism. We believe in the second evidence (dimorphism); as both forms have the same number of ribs, and the attachment area is equal, in addition to the overall shell characters of the genus *Rhynchostreon* Videt and Platel (2003).

Cluster Analysis

In the next step, we applied cluster anlaysis in order to evaluate the morphological similarities between Exogyrinae oysters. In cluster analysis we search for patterns in a data set by grouping the (multivariate) observations into clusters. The goal is to find an optimal grouping for which the observations or objects within each cluster are similar, but the clusters are dissimilar to each other. We hope to find the natural groupings in the data, groupings that make sense to the researcher. Cluster analysis has also been referred to as classification, pattern recognition (specifically, unsupervised learning), and numerical taxonomy. The techniques of cluster analysis have been extensively applied to data in many fields, such as medicine, psychiatry, sociology, criminology, anthropology, archaeology, geology, geography, remote sensing, market research, economics, and engineering. A common distance function is the Euclidean distance between two vectors, $x = (x_1, x_2, ...x_p)$, and $y = (y_1, y_2, ...y_p)$, which is defined as:

$$d(\mathbf{x},\mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})'(\mathbf{x} - \mathbf{y})} = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2}$$

The combined tree provides better resolution of overall phylogeny (Fig. 6.11). At similarity equals 0.925 the major tree splits into four groups containing all genera and subgenera, these lineages have a distinct similarity distance and provide primary phylogenetic tree. The main problem is to estimate the time boundaries between species occurrences, which is absolutely undetermined by the cluster analysis, therefore we applied cladistic analysis in the next step.

Cladistic Analysis

We used cladistic analysis to assign time boundaries for the examined oyster members. Cladisitc analysis is a semi-objective analysis of relationships between taxa from morphological or genetic evidence; the analysis searches the data matrix for the most parsimonious tree or trees. The results are presented in Fig. 6.12.

The first taxon, *Exogyra plexa*, is treated as outgroup, and is placed at the root of the tree. The topology of the tree indicates that substantial revision of the forms belonging to the genus *IIymatogyra* is required. All the forms attributed to the genus do not lie on a single time unit or clade. *I.* (*A.*) *Africana* (morphotypes A and B) lies between different clades, indicating change in the environmental conditions with time. The genus *Rhynchostreon* represents forms (*suborbiculatum* morphotypes A and B) that are linked within a single clade or time unit (Fig. 6.12).

Matching with Stratigraphy

Since some characters are ordered with respect to the states exhibited by the earliest taxa, and *E. plexa*, the earliest species of Exogyrinae, is used as outgroup,
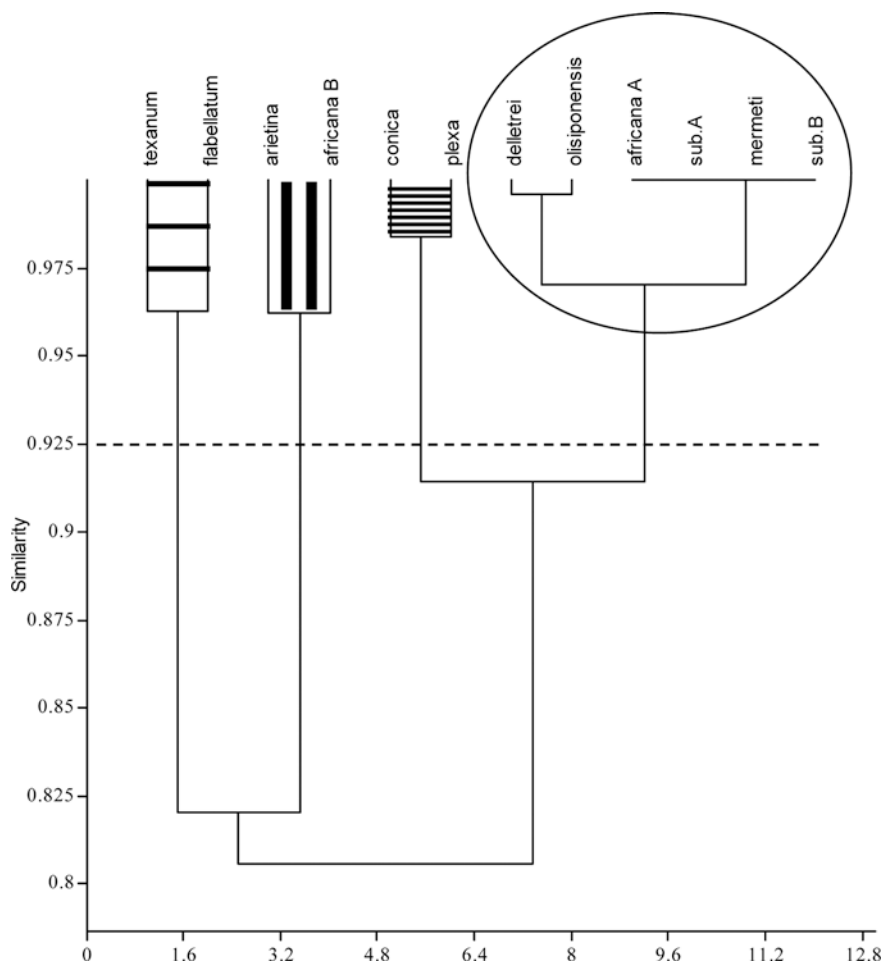
Fig. 6.11 Dendrogram obtained from Euclidean based clustering, after grouping the data matrix rows

it is therefore debatable whether there is quite good congruence between the cladogram and the stratigraphical occurrence of the fossils. The occurrence of *E. plexa* and *E. conica* at the base of the clade pattern agrees with their earlier stratigraphical occurrences at the late Albian/early Cenomanian interval. For the other cases, the cladogram has located *C. flabellatum* (late Cenomanian) at earlier stage than *C. texanum* (middle Albian). Furthermore, the cladogram placed *E. (C.) olisiponensis* and *Gyrostrea delletrei* (both of late Cenomanian age) at different levels.
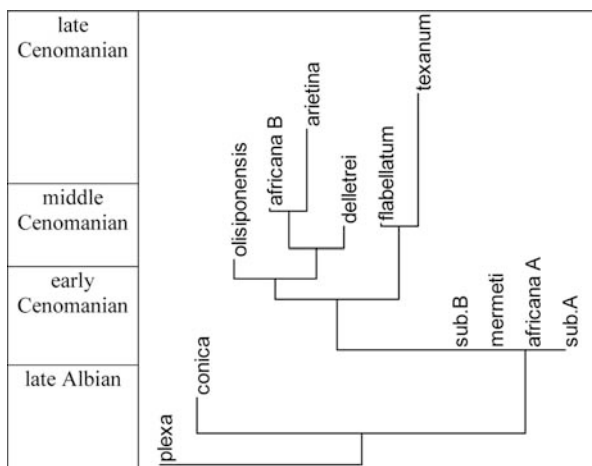
Fig. 6.12 Phylogram based on cladistic analysis

Correspondence Analysis

Since we could identify the morphological similarities of oyster members, subsequently we have to shade the light on the paleoecological factors that affected the species occurrences and richness, using correspondence analysis (CA).

Correspondence analysis is a graphical technique that represents the information in a two-way contingency table, which contains the counts (frequencies) of items for a cross-classification of two categorical variables. Using correspondence analysis, we can construct a plot that shows the interaction of the two categorical variables along with the relationship of the rows to each other, and of the columns to each other.

The algorithm of Davis (1986) is herein considered. The routine finds the eigenvalues and eigenvectors for a matrix containing the chi-squared distances between all data points. The eigenvalues, giving a measure of the similarity accounted for by the corresponding eigenvectors, are given for the first four most important eigenvectors (or fewer if there are less than four variables). The percentages of similarity accounted for by these components are also given. The main target of the Correspondence analysis is to reduce and interpret large multivariate ecological data sets with environmental or other gradients.

In correspondence analysis, we plot a point for each row and a point for each column of the contingency table. These points are, in effect, projections of the rows and columns of the contingency table onto a two-dimensional Euclidean space. The goal is to preserve as far as possible the relationship of the rows (or columns) to each other in a two-dimensional space. If two row points are close together, the profiles

of the two rows (across the columns) are similar. Likewise, two column points that are close together represent columns with similar profiles across the rows. If the two variables are denoted by x and y, then the assumption of independence can be expressed in terms of probabilities as:

$$\mathbf{P(xi\ y\ j) = P(xi)P(y\ j), i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b,}$$

Where xi and yj correspond to the *i*th row and *j*th column of the contingency table, we can estimate:

$$\mathbf{pi\ j = pi.\ p.\ j, i = 1, 2, \ldots, a;\ j = 1, 2, \ldots, b}$$

Correspondence analysis and its plots suggest some important criteria on assessing morphological characters variation, with changes in the environmental conditions, within the different oyster species. From Table 6.1, it is evident that the first and second eigenvalues account for more than 96% of the trace. Consequently, Fig. 6.13 represents the scores of the 1st versus 2nd correspondences. The results are same as for cluster analysis, however environmental conditions can be predicted from correspondence analysis. The projection of the genus *Rhynchostreon* with its species and forms on a single point (Fig. 6.13) indicates the identity of the favorable conditions for all forms which occur on the same stratigraphical horizon (early Cenomanian), with equal response to the ecological factors, as the overall shell shape and size (L/W & L/B).

Figure 6.13 indicates that species of the genus *Ilymatogyra*, *I. arietina* and *I. (A.) Africana* morphotype B, are grouped according to the change in Theta1, the angle between the umbo and the tangent for the postero-ventral margin. Whereas, *I. (A.) Africana* morphotype A is located further away from the genus *Rhynchostreon*; suggesting the occurrence of an intermediate form between the two genera, with shell shape and ecological conditions related to the *Rhynchostreon*, and diagnostic features related to the genus *Ilymatogyra*. It is also observable, from the same figure, that Theta3, the angle between postero-ventral margin and the equator of the shell, assembles *C. texanum* and *C. flabellatum* within the same genus.

Table 6.1 Summary of correspondence analysis

Axis	Eigenvalue	% of total
1	0.0554519	58.52
2	0.0361253	38.124
3	0.00279102	2.9455
4	0.000358798	0.37865
5	E-052.68978	0.028386
6	E-062.62578	0.0027711
7	E-071.73952	0.00018358
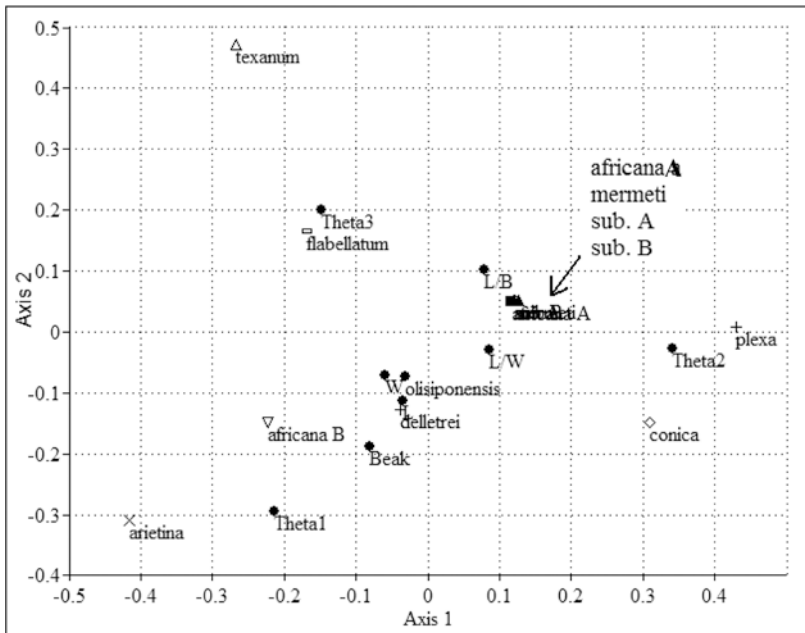
Fig. 6.13   Plot of points representing the 10 species on rows and the 8 variables on columns

E. plexa and *E. Conica* belong to each other by the angle Theta2, the angle between equator and the lower limit of the umbo. *E. (C.) olisiponensis* and *G. delletrei* are linked together because of the closeness of their shell diameters (L &W), while the other taxa are linked together based on the L/W & L/B ratios.

We can also conclude that Theta3 increases with the decrease of the valve periphery, which indicates a great convexity. This convex shape of the valve permits the epifauna to fix their bodies on the substrate more easily. In general, species with high convexity (*Exogyra*) are almost found on sandy substratum, while those with high Theta3 or large periphery accumulate with silt or mud substratum (*Ceratostreon*) (see Videt 2003).

The Proposed Phylogeny

Based on the collected material, and the literatures, we could propose the phylogeny shown in Fig. 6.14. Left valve illustrates some of the trends exhibited in the evolution of the *E. plexa* to *E. conica*. The most obvious trend is toward a larger beak with a greater number of turns, and a greater pitch. The overall size of the valve increases as well. Furthermore, the evolution of the *E. plexa* to *E. conica* is represented by the trend towards increasing the folds and the ribs. *E. plexa* possesses ribs near the tip of the beak. In addition, there is a trend towards reducing the attachment area. For
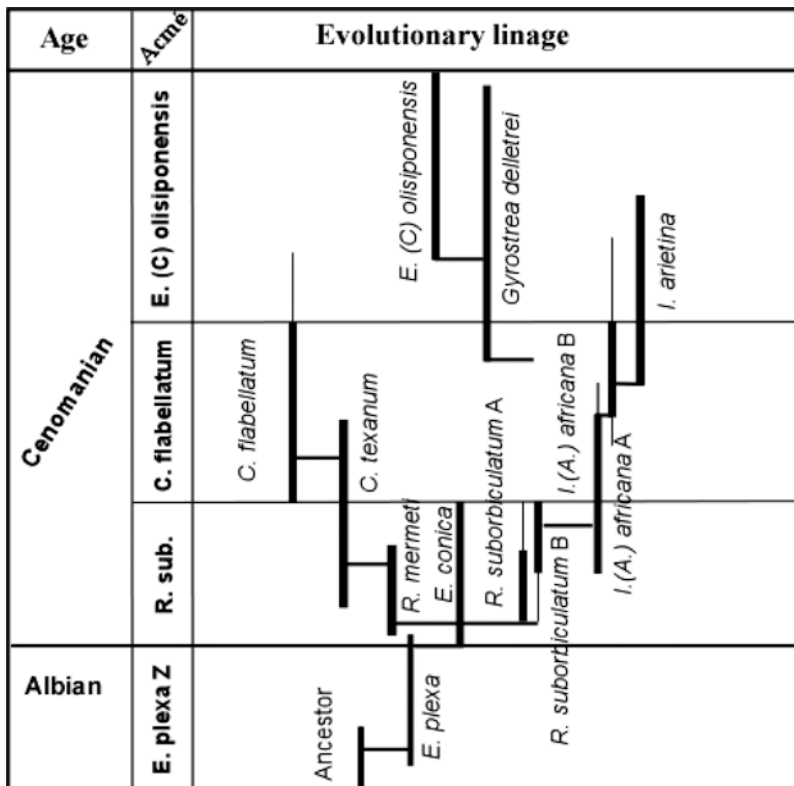
Fig. 6.14 Proposed phylogenetic tree in the Albian – Cenomanian Exogyrinae oysters

specimens of *E. plexa*, the attachment scar is often comparable to the valve in size. For *E. conica*, the attachment area is slightly absent.

We can observe the same evidence in the evolution from *I. arietina* to *I.* (*A.*) *Africana* (morphotypes A and B). There is a trend towards increasing the size with increasing the valve convexity, valve turns, and beak area. Another trend towards increasing the lower valve convexity and beak area can be shown in the evolution from *E. conica* to *R. suborbiculatum* morphotypes A and B, and *R. mermeti*. On the other hand, the ribs and valve turns decrease in the same path.

A good gradational transformation of *Ceratostreon texanum* to *C. flabellatum* is evident. This transformation is shown by increasing the size and surface ornament (ribs and nodes). The curvature of the beak is also increased with the decreasing of the valve turns. *C. flabellatum* is characterized by more conspicuous nodes than *C. texanum*, especially at the ventral margin. The beak turns also decrease through the same path.

Overall, the present study indicates that careful examination of microstructure and species from different stratigraphical horizons may reveal further details about systematics and phylogeny of the Upper Cretaceous oysters.

The Evolutionary Models

Modern theories of species evolution are determined by two famous models that reflect emphasis on different aspects of the evolutionary process. The first model is "Phyletic gradualism" that was originated with Darwin's concept of organic evolution. The model holds that change in frequency of morphological characters occurs, gradually, in response to the selection pressure on the species populations. Abrupt change in morphology, in the fossil record, is viewed as something different from the speciation event (such as sampling, preservation failures or migration). If a complete record is preserved it would be expected to show continuous gradual morphological change from one species to the other.

The second model is called "punctuated equilibrium"; it characterizes species as real natural entities that are dominated, for most of their history, by morphological stability, and are characterized by abrupt origins and termination in space
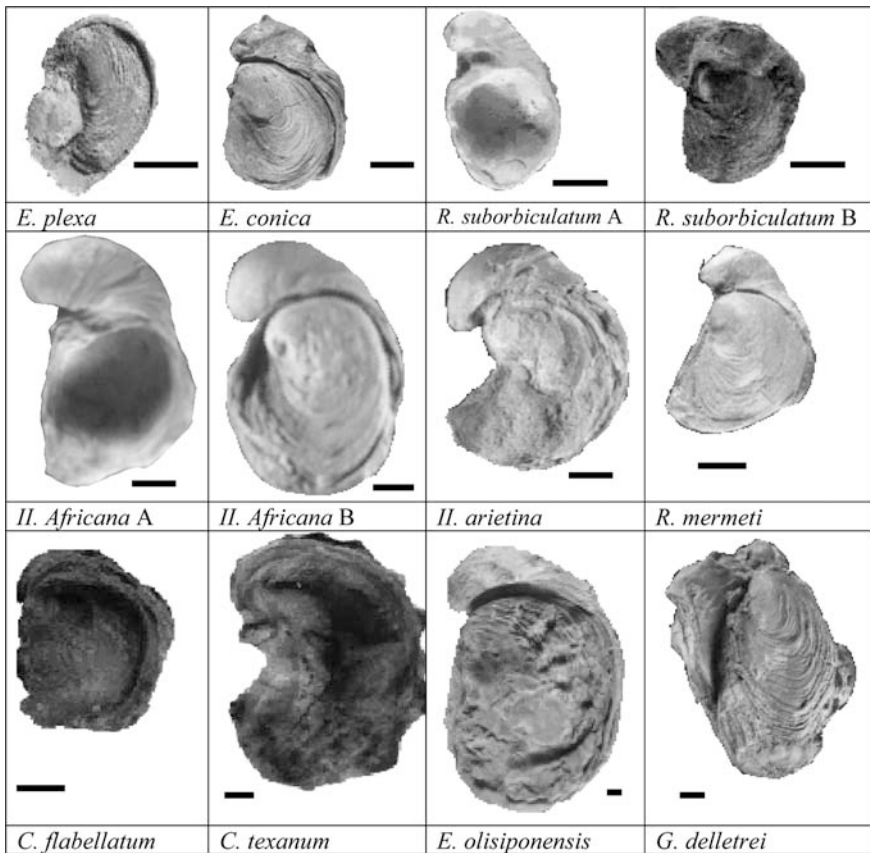
Fig. 6.15 Representatives of the examined oyster species

and time. Speciation is accomplished by the relatively rapid (perhaps 5,000–50,000 years as order of magnitude) development of reproductive isolation in some geographically isolated population of the parent species, followed by longer period of stasis in the daughter populations. Zonal boundaries for pyloric gradualism will be inherently imprecise, and errors will be introduced in attempted time correlation as a result of the difficulty for definition of species limits. On the other hand, punctuated equilibrium allows easily recognition of zonal boundaries precisely.

The investigated material (Fig. 6.15) suggests modification for the above mentioned models. It is believed that speciation goes flourished onward until reaching a maximum density, and then it is gradually vanished, due to the environmental
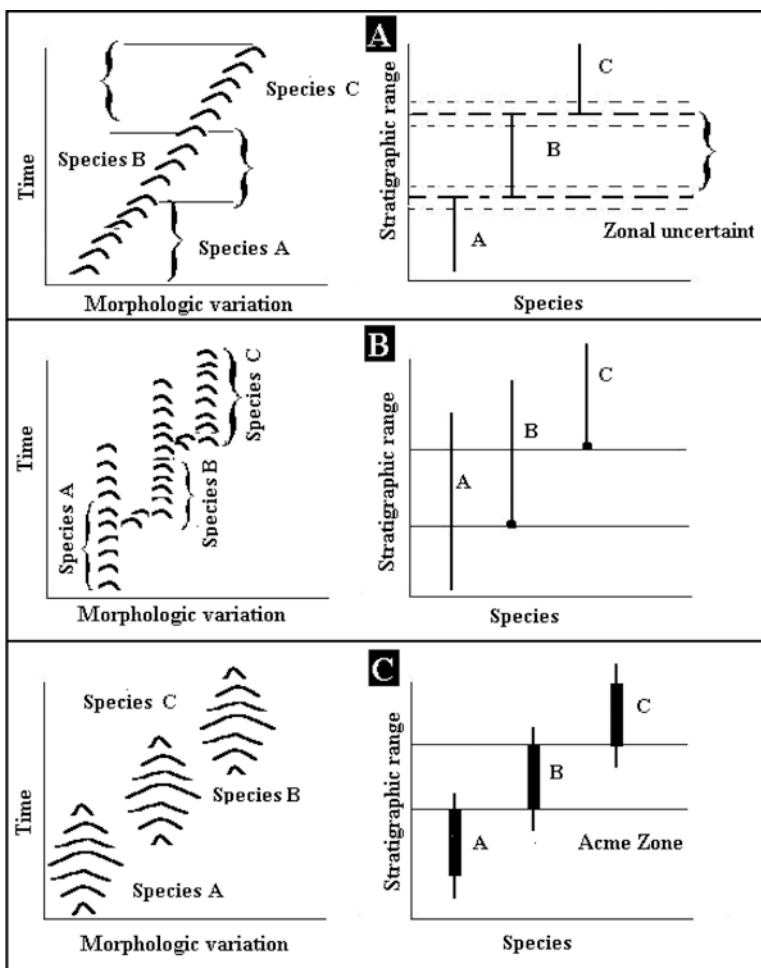
Fig. 6.16 Implication of species evolution models to the definition of biochronological zones. Phyletic gradualism (a), Punctuated equilibrium (b), Proposed model (c)

changes, with the appearance of gradational morphotypes of the same species. The stages of these morphotypes are almost of short duration. The morphotypes rise up to a new species that starts with low density, then its density increases onward, and repeats the cycle of the parent. Zonal boundaries according to this new model can easily be determined (Fig. 6.16).

In conclusion, we can ensure that the faunal associations, which are revealed by newly discovered, well preserved, and therefore, highly diverse assemblages, allowed the recognition of some phylogenetic trends of oysters. The vertical distribution of the identified taxa reveals six acme zones of the Upper Creatceous age. The present study demonstrated that traditional morphometrics could provide a useful tool in the study of the oyster morpho-evolutionary changes.

References

Abdelhady, A. A. (2007) Stratigraphical and Paleontological Studies on the Upper Cretaceous Strata, North Wadi Qena, Eastern Desert Egypt. M.Sc. Thesis, Minia University, 173 p.

Aqrabawi, M. (1993) Oysters (Bivalvia-Pteriomorphia) of the Upper Cretaceous rocks of Jordan. Palaeontology, stratigraphy and comparison with the Upper Cretaceous oysters of northwest Europe. Mitteilungen des Geologisch-Palaontologischen Institutes der Univ. Hamburg 75, 135 p.

Böse, E. (1919) On a New *Exogyra* from the Del Rio Clay and Some Observations on the Evolution of *Exogyra* in the Texas Cretaceous, Bulletin No. 1902, University of Texas, Bureau of Economic Geology.

Davis, J. C. (1986) Statistics and data analysis in geology. John Wiley & Sons, New York.

Dhondt, A. V. (1992) Palaeogeographic distribution of Cretaceous Tethyan non-rudist bivalves. In New aspects on tethyan cretaceous fossil assemblages 9. Schriftenreihe Erdwisscntschaftliche Kommission der sterreichische Akademie der Wissenschaften, 75–94 pp.

Dhondt, A. V., and Jaillard, E. (2005) Cretaceous bivalves from Ecuador and northern Peru. Journal of South American Earth Sciences 19, 325–342.

Haq, B., Hardenbol, J., and Vail, P. R., (1987) Chronology of fluctuating sea levels since the Triassic. Science 235, 1156–1167.

Hautmann, M. (2001) Taxonomy and phylogeny of cementing Triassic bivalves (families Prospondylidae, Plicatulidae, Dimyidae and Ostreidae). Palaeontology 44, 339–373.

Malchus, N. (1990) Revision der Kreide-Austern (Bivalvia: Pteriomorphia) Egyptens (Biostratigraphie, Systematik), Berliner geowiss. (A) 125, 194.

Malchus, N. (1996) Palaeobiogeography of Cretaceous oysters (Bi-valvia) in the western Tethys. Mitteilungen aus dem Geologisch-Palä¨ontologischen Institut der Univ. Hamburg 77, 165–181 pp.

Mancini, E. A. (1978) Origin of the grayson micromorph fauna (upper cretaceous) of north-central Texas. Journal of Paleontology 52, 1294–1314.

Márquez-Aliaga, A., Jiménez-Jiménez, P. A., Checa, A. G., and Hagdorn, H. (2005) Early oysters and their supposed Permian ancestors. Palaeogeography, Palaeoclimatology, Palaeoecology.

Moore, R. C. (1969) Treatise on invertebrate paleontology. Part N volume (1–3) Mollusca 6, Bivalvia. Geological Society of America and University of Kansas Press, 1299 p.

Oyvind, H., David A. T., and Paul R. D. (2001) Past: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica, 4(1), art. 4–9.

Rohlf, F. J. (1998) TPSdig, version 1.20. N.Y.: State University at Stony Brook. (program).

Seeling, J., and Bengston, P. (1999) Cenomanian oysters from the Sergipe Basin, Brazil. Cretaceous Research 20, 747–765.

Smith, A. B., and Bengtson, P. (1991) Cretaceous echinoids from north-eastern Brazil. Fossils and Strata 31, 88 pp.

Videt, B. (2003) Dynamique des paleoenvironnements a huitres du cretace superieur nordaquitain(so france) et du mio-pliocene andalou (se espagne): biodiversite, nalyse sequentielle, biogeochimie., Ph.D., universite de rennes, France, 1–304 pp.

Videt, B., and Platel, J. P. (2003) Les ostréidés des faciès lignitifères du Crétacé moyen du Sud-Ouest de la France (Charentes et Sarladais) C. R. Pale 4, 67–176 pp.

Waller, T. R. (1978) Morphology, morphoclines and a new classifi-cation of the Pteriomorphia In Walter, S., Herrmann, A., & Bengston, P. Stratigraphy and facies analysis of the Cenomanian-Turonian boundary suc-cession in the Japaratuba area, Sergipe Basin, Brazilian Journal of SAESB 284, 345–365.

Yonge, C. M. (1978) On the Dimyidae (Mollusca, Bivalvia) with special reference to Dimya corrugata Hedley and Basiliomya goreaui Bayer. Journal of Molluscan Studies 44, 357–375.

Yonge, C. M. (1980) On the dimyidae and plicatulidae—proposed superfamily plicatulacea. Journal of Malacology Society of Australia 4, 241– 242.

Yurke, and Charles (2006) Evolution of the exogyrinae oyster. Internal Report, University of Texas.

Arctic Council, 2004. Impacts of a warming climate: arctic climate impact assessment. Cambridge University Press, Cambridge.

Part III
Applications (Chaps. 7 to 15)

Chapter 7
Combining Shape Data and Traditional Measurements with the 2B-PLS: Testing the Covariation Between Avian Brain Size and Cranial Shape Variation as an Example

Jesús Marugán-Lobón

Idea and Aims

The Two-block Partial Least Squares (2B-PLS) is a multivariate exploratory tool of geometric morphometrics intended to test the covariation between two sources of quantitative data. The method allows testing complex hypotheses of biological association between shape and any quantifiable item of the natural world, not only between shape data. However, this latter ability has been scarcely exploited so far despite its great utility. This chapter exemplifies how the 2BPLS can combine traditional measurements (herein, measurements of mass) with Procrustes shape data, and test their covariation. The example is in the context of avian morphological evolution, and it is tested whether the differences in braincase shape between avian taxa are or are not associated with avian brain size variation. In this particular context obtaining landmark data of the avian cranium is easier than collecting information about brain variation. One parallel benefit of using the 2B-PLS in combination with traditional data is that the latter abounds in literature, and this helps testing complex hypotheses promptly. The limitation of this approach is the obliged need of a reasonable pair-wise species overlapping between the two sources of information. In the present case the 2B-PLS could not find any statistically significant pattern of covariation between brain masses and the shape variation of the avian cranium. This suggests that, as in mammals, avian cranial morphological diversity is a complex phenomenon, which thereby needs to be further investigated.

Introduction

Morphometrics have been an essential tool to discover processes of the biological world (Rohlf 1990; Roth and Mercer 2002), and landmark-based Geometric Morphometrics (GM) is among the newest additions to this operational field. The

J. Marugán-Lobón (✉)

Unidad de Paleontología, Dpto. Biología, Universidad Autónoma de Madrid, 28049, Cantoblanco, Madrid, Spain

e-mail: jesus.marugan@uam.es

advent of GM was received as a revolution in the last decade of the past century (Rohlf and Marcus 1993, Adams et al. 2004), and today its core and its principal toolkit are very much developed and standardized. However, some of its features are still unexploited to their full potential. This is for instance the case of the ability of the Two-block Partial Least Squares (2B-PLS; Rohlf and Corti 2000) to test the interplay between shape data and traditional metrics, to which this chapter is devoted.

The 2B-PLS is a method useful for exploring patterns of relationship (statistical covariation) between one set of shape variables and another set of quantitative variables of any sort. The ability to test the correlation between two sets of shape variables has been the most widely exploited side of the technique, principally to test hypotheses of Morphological Integration, a complex biological phenomenon that implies the coordinated change between parts of organisms in the course of development or evolution (Olson and Miller 1958). However, the 2B-PLS is also capable of testing the interplay between shape variables and another set variables which can be of any quantitative sort. This analytical tool thus allows testing the relationship between shape variation and any measurable item of the natural world, under the umbrella of a reasonable biological hypothesis that justifies their association. However, the potentials of this ability remain underutilized today. Adams and Rohlf (2000) successfully carried out a study of ecological character displacement in which the technique was used to test the degree of association between cranial morphology (shape) and food resources in salamanders which were computed as food frequency values. Marugán-Lobón and Buscalioni (2006) explored craniofacial shape evolution among avians comparing it to angular differences referenced to functional head posture. More recently, Windhager et al. (2008) in an awe-inspiring work exploited this side of the 2B-PLS in a psychological enquiry, by exploring the possible influence of car "faces" (their shape) upon the subconscious in human perception. The non-shape data was achieved coding the responses to questionnaires. So far, however, no study has tested the relationship between shape and the most common biometric proxies to size, such as lengths, or masses. This may be due to the fact that centroid size, the estimation of scale in geometric morphometrics, has several advantageous properties than the traditional estimates of size (Bookstein 1991). However, there can be cases in which using this approach may be useful, such as in the case which is exemplified in this study.

The Example; the Potential Involvement of Avian Brain Size Variation in Avian Cranial Shape Diversity

Understanding the proximate causes of cranial shape diversity, not only in birds but in any vertebrate, is an important research motto since the way in which the cranium is organized has critical evolutionary implications. The structural differences of the cranium decisively influence a variety of head and body postures (Duijm 1951; Marugán-Lobón and Buscalioni 2006), and head postures are implicated in the variety of feeding behaviors and locomotion patterns displayed by each species

(i.e., the way in which each species feeds, hunts, walks, runs, flies, etc.; Witmer et al. 2003). So, what major features are these which characterize the cranial differences between birds?

The cranium of any vertebrate is a very complex system, and it would be too long to explain the anatomical details of its architectural variation. What is perhaps more intuitive to bear in mind is that most of these entwine in one way or another with the way in which the neck attaches the head. Let us use the human cranium as an example. In humans the neck stands vertical to the ground instead whereas in most mammals the neck is horizontal, and thus, our medulla enters the head ventrally through the foramen magnum (this is a large hole which serves as a path to the medulla from the cranial cavity down to the vertebral column; Fig. 7.1). This anthropoid feature is ubiquitous yet today how it was achieved remains an intriguing issue. Physically, it involves a series of complex anatomical re-arrangements which make the cranial base (the floor underneath the brain) to organize differently from the "normal" mammalian plan, and this seems to primarily happen within the midline, and not on the lateral sides, of the cranium (Enlow and Hans 1996). The close vicinity between the cranial base and the brain, on the other hand, clearly suggests that the marked enlargement of the primate brain is involved in altering these topological differences of the cranium (Lieberman et al. 2008).

The above explanation was useful to get a first picture of the idea dealt with herein before speaking of birds. In effect, this was useful since we are familiar with the shape of our head, and not of birds. What is less familiar to us is that there are certain bird species in which the cranium has the foramen magnum opening ventrally instead of caudally, and therefore, their neck attaches to the head vertically like in primates. Examples are swifts, hawks, and owls, although there are many others (Marugán-Lobón and Buscalioni 2009). In other avian species, such as cormorants, chicken, or grebes the neck is connected with the head like in most
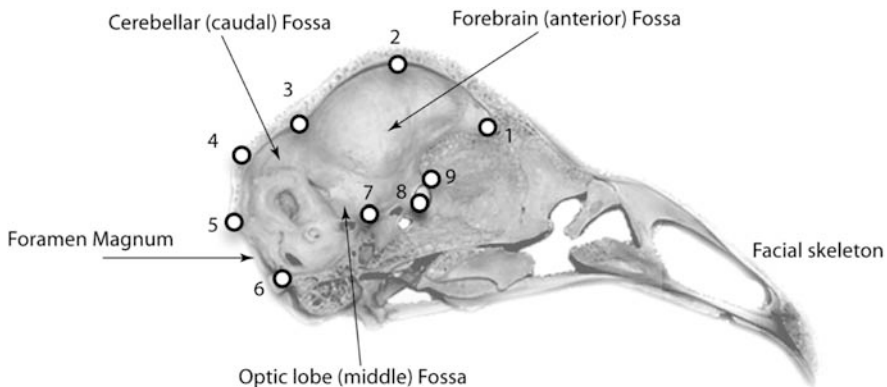
Fig. 7.1 Internal view of the skull of a bird. An example of the configuration of landmarks (1–9) is superimposed over the cranium. Notice how the endocranium reflects the tight fitting of the brain therein (one can readily follow the external and internal contours of the brain; follow arrows and labels as a guideline). The cranial base extends between landmarks 1, 6, 7, 8 and 9

reptiles and dinosaurs, horizontally instead of ventrally. The cranial base in birds (see Fig. 7.1 caption) is involved in these head-neck attachments as well, thus the question is do these cranial differences between birds depend on brain size as they do in mammals? From the early Twentieth century up to its midst, some authors explained that this could be the case whereas others defended that it wouldn't (Zusi 1993). However, none of these morphologists ever tested the relationship between brain size and cranial shape on empirical bases, and the debate just faded away until our days.

This study will explore if cranial shape variation between birds (or, if its side-effect, the alternation of neck-head attachment) is associated with their brain size variation, as it is proposed in mammals. If this would be the case, the 2B-PLS should be able to find a statistically significant trend of covariation between the two systems such that the larger their avian brain gets, the more ventrally oriented the foramen magnum would become (i.e., the more vertical the attachment of the neck), just as the theory explains how this cranial organization is achieved by primates (Ross et al. 2004).

Materials and Methods

Avian Brain Sizes

Acquiring real brain data of the avian brain is elaborate since there are no collections available of brains as they are of crania. The sampling effort that one needs to make in order to collect brains from birds as different as hawks, cormorants, owls, swifts, or ostriches, most of which live in the wild, is thus too large (and needs some expertise) for a work which in fact, is not about encephalization *per se*. However, there are some works, such as Portmann (1946, 1947) in which real brain masses were studied and are now available. This source is a classic, and it has been used in textbooks and neurologic research as a reference ever since it was first published (see e.g.; Pearson 1972; Harvey and Krebs 1990; Striedter 2005). The original paper contains data of the weighted masses of the most distinguishable parts of the gross morphology of the central nervous system. These are the brainstem plus a small part of the diencephalon (this is altogether known as the encephalic trunk, and will be herein called the medulla), the cerebellum, the optic lobes and the telencephalon (the forebrain) (Fig. 7.2a). At the same time the sampling of this work covered a reasonably broad taxonomic spectrum of birds. However, the author came up with what he called a "basal cipher", a denominator used to calculate and compare what he called indexes of encephalization. This cipher represented the mass of the brainstem rest of an arbitrary galliform bird of the equivalent weight (body mass) of the species whose brain was being compared. In order to recover the original mass values, here each of the encephalisation indexes for each brain part was multiplied by the basal cipher which was provided by the author. It is also noteworthy that these measurements represent the masses of the most differentiable parts of the gross morphology of the
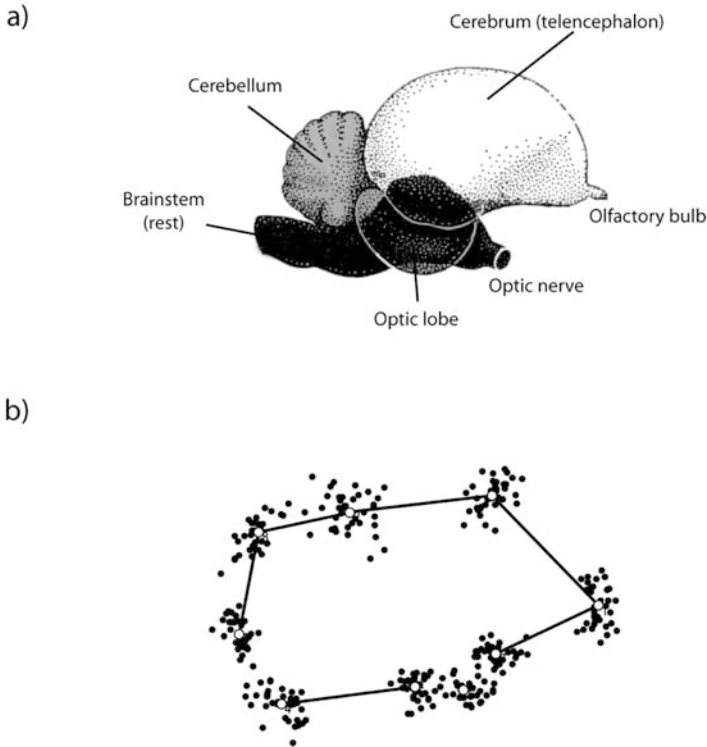
Fig. 7.2 (**a**) Coordinates of the Procrustes residuals of the configurations of landmarks superimposed over the consensus which the 2B-PLS will attempt to correlate with brain masses. (**b**) Scheme of the gross morphology of the avian brain modified from Portmann (1946)

brain instead of the size of the brain as a whole. This is important since it would be possible that the size of each brain part varies somewhat independently from the rest (Striedter 2005), and this could have different implications upon cranial shape than if the whole brain changes its size with all its major parts doing so at unison. The 2B-PLS is capable of detecting whether this may or may not happen (see below).

Cranial Shape Data

Collecting landmark data from the avian cranium is trickier than in mammals since cranial sutures are lost in late ontogeny. Furthermore, it is not easy to find avian skull collections in which skulls are opened so to acquire data from the endocranium. Endocranial data is needed since it is therein, and in the midline, where the most important features of cranial base variation are located (those involved in differential neck attachment) (Fig. 7.1). While today this is becoming feasible with CTscans, this is still a bit complicated and expensive method. However, Marugán-Lobón and Buscalioni (2009) were able to collect this information from an old and

beautiful collection which was hand-prepared by Hesse (1907), and which is housed in the Museum für Naturkunde in Berlin, Germany. This sample containing land-mark data was re-used herein, and is describe a bit below. However, the landmarks only represent coordinates of the cranial base in the midline (most other sites within the avian endocranium are featureless, made up by smooth surfaces, which makes it difficult to register landmarks). Being forced to registering the landmarks in the midline forces this data to be in two dimensions, and this impedes estimating the size of the whole (3D) endocranial cavity with centroid size (the square root of the sum of the squared distances between the landmarks; Bookstein 1991), which would be equivalent to estimating the size of the whole brain (Iwaniuk and Nelson 2002). This is also a reason why instead, real brain sizes are needed for this study.

When using literature resources to test a hypothesis of covariation with the 2B-PLS is that the digitized specimens (for landmark acquisition and shape data) need to overlap with the species whose measurements are available in the literature. Otherwise the comparisons would not make sense. The original brain data from the literature source (Portmann 1946, 1947) overlapped with $n = 37$ species out of 72 specimens which were available from in the original landmark data-base (Marugán-Lobón and Buscalioni 2009). This resulting subsample is somewhat small in number of individuals, yet it contains a significantly ample representation of birds Orders (16) after Livezey and Zusi (2007). For instance, the sample includes representatives of the Paleognathae, which includes the ostrich and allies, the Galloanserae, com-prised by the galliforms (chicken and related forms) and the anseriforms (swans and ducks), plus several clades within Neoaves, which encompasses all the rest of imaginable birds such as passerines, parrots, swifts, cormorants, waders etc. (Table 7.1).

The locations of nine homologous landmarks were registered in digital pictures of each specimen (Fig. 7.1, and Table 7.2), using F. J. Rohlf's TPSdig (v. 1.40), and the shape data (the Procrustes residuals, Fig. 7.2b) was thereafter obtained using the Generalized Least Squares procedure as described by Gower (1975) and (Rohlf and Slice 1990). This shape information is what remains after translating, rotating, and scaling all configurations to a common coordinates system -the Procustes average, or consensus; Bookstein 1991; Fig. 7.2b).

The Partial Least Squares Method

The Two-block Partial Least squares (2B-PLS; Rohlf and Corti 2000) will be used to explore the covariation between the shape variables and brain masses. Rohlf and Corti (op. cit.) disclose its matrix algebra and discuss several of its uses, and emphasize its differences with Canonical Correlations. Zelditch et al. (2004; p. 261) detail other potential uses of the technique, as well as its major differences with Multiple Regression and Principal Components analyses (see the method in; Lattin 2002). In a nutshell, the 2B-PLS is a multivariate technique that seeks for the covariation among the observations in terms of a few dimensions that are linear combinations of the original variables. These dimensions are inter-block generated,

Table 7.1 Species names, orders, and institutional labels

Species name	Order	Signature	Species name	Signature	Order
Strutio camellus	Strutioniformes	AMNH 4261	*Buteo buteo*	ZMB 149	Falconiformes
Cygnus odor	Anseriformes	ZMB 986	*Falco tinnunculus*	ZMB 765	Falconiformes
Coturnis coturnis	Galliformes	ZMB 308	*Accipiter nisus*	ZMB 1	Falconiformes
Pavo cristatus	Galliformes	ZMB 902	*Asio flammeus*	ZMB 334	Strigiformes
Perdix perdix	Galliformes	ZMB 310	*Asio otus*	ZMB 768	Strigiformes
Gallus gallus	Galliformes	ZMB 77	*Anthropoides virgo*	ZMB 853	Gruiformes
Tetrao urogallus	Galliformes	ZMB 24	*Picus viridis*	ZMB 505	Piciformes
Ardea cinerea	Ardeiformes	ZMB 396	*Podiceps cristatus*	ZMB 803	Podicipediformes
Larus canus	Charadriiformes	ZMB 774	*Amazona automnalis*	ZMB 778	Psittaciformes
Scolopax rusticola	Charadriiformes	ZMB 160	*Fulica atra*	ZMB 489	Ralliformes
Numenius arquata	Charadriiformes	ZMB 45	*Rallus aquaticus*	ZMB 894	Ralliformes
Haematopus ostralegus	Charadriiformes	ZMB 968	*Porphyrio posphyrio*	ZMB 646	Ralliformes
Vanellus vanellus	Charadriiformes	ZMB 887	*Lanius exultor*	ZMB 674	Passeriformes
Phoenicopterus ruber	Ciconiiformes	ZMB 809	*Corvus corone*	ZMB 148	Passeriformes
Ciconia ciconia	Ciconiiformes	ZMB 253	*Parus sp.*	ZMB 762	Passeriformes
Columba sp	Columbiformes	ZMB 248	*Pica pica*	ZMB 623	Passeriformes
Goura cristata	Columbiformes	ZMB 879	*Pyrrocorax pyrrhocorax*	ZMB 697	Passeriformes
Cuculus canorus	Cuculiformes	ZMB 242			
Phalacrocorax carbo	Pelecaniformes	ZMB 570			
Pelecanus onocrotalus	Pelecaniformes	ZMB 763			

Table 7.2 Description of landmarks

Number	Description
1	Olfactory nerve
2	Mid-point between LM 1 and 3 at the edge of frontal
3	Ventral edge of crest separating forebrain and cerebellar fossae
4	External junction between supraoccipital and parietal bones
5	Dorsal rim of foramen magnum
6	Ventral rim of foramen magnum
7	Dorsal-most point of Sella
8	Ventral edge of the optic foramen
9	Dorsal edge of the optic foramen

its components are ordered according to covariance, and the variance-covariance matrix does not need to be squared, which is mathematically accomplished by using a single value decomposition (SVD) of the matrix (the SVD is explained in; Lattin et al. 2002). This is why the technique in GM is also known as Singular Warp Analysis (Bookstein et al. 2003). Each dimension of the SVD, so-called singular values, associates to a pair of vectors (one vector per block of variables, where each is a linear combination of the coefficients of the each block, and each is called a *singular axis*), and the explained covariance is measured when pairing the two axes of each dimension. In other words, the technique will seek for the latent relationships between the variables within each block (i.e., search for the variables within each block that are maximally related), and construe linear combinations of them that account for as most covariation as possible between the two blocks. In principle, it is only necessary to standardize the non-shape data if the variables are in different units (Rohlf and Corti 2000). The 2B-PLS analysis was been performed with the TPSpls module (v1.18) programmed by F. J. Rohlf, which, like TPSdig, is available from the server of the Department of Ecology and Evolution, SUNNY at Stony Brook, US, at http://life.bio.sunysb.edu/morph/.

Results

The first singular value of the 2B-PLS analysis can explain as much as 91.31% of the total covariance between brain weights and cranial shape data. However, neither this covariation nor the correlation between its pair of vectors is statistically significant. Permutations tests estimate that the covariance value can be equalled or exceeded randomly (74.33% of the samples out of 1,000 permutations did so), and so is the case for the correlation between its vector pairs, which is low ($r = 0.43$; see highlighted scatter-plot in Fig. 7.3), and permutations tests estimate that this correlation value can be equalled or exceeded randomly (11.69% of the samples out of 1,000). Thus, there is not sufficient evidence to state that absolute brain size associates with the changes in the topology of the cranium which alter the attachment of the head to the neck between birds.
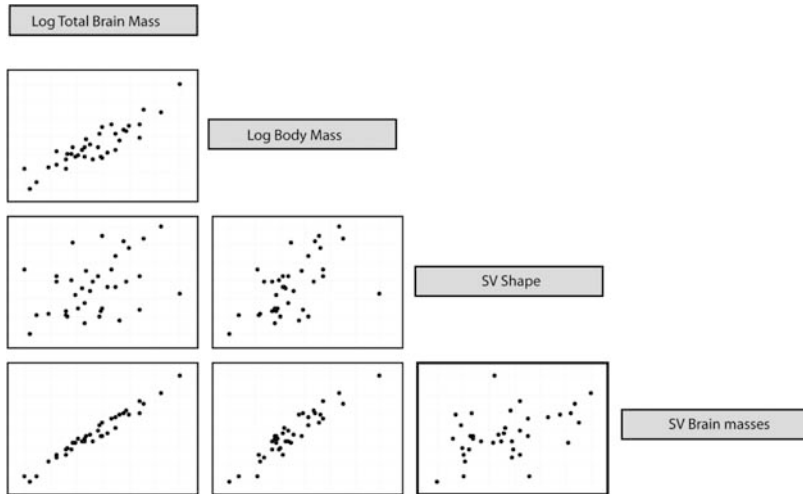
Fig. 7.3 Matrix of the relationships between log-body mass, log-centroid size, and the two vectors of the first singular value obtained by the 2B-PLS. The relationship between the vectors is the highlighted box at bottom of the matrix (SV stands for the vector of the singular value). Notice the marked linear relationship between both size variables (body mass and total brain mass) and the SV-vector of brain masses

Discussion

The Two-block Partial Least Squares is a technique which was implemented into the statistical toolkit of geometric morphometrics as a method to explore the interplay between shape and other sources of quantitative data. However, most studies are often focused in using the technique to test the covariation between shapes, and its ability to test the covariation between shape data and traditional variables remains quite unexploited. The example used in the present chapter was meant to explore the relationship between real measurements of brain masses (the mass of the neural system) and cranial shape (a bony system). The question asked was whether the size of the brain has any role in shaping the varied architectural plan of the avian cranium. The results of the 2B-PLS have yielded a non-statistically significant first singular value between brain size and cranial shape among birds, entailing that the expected covariation is rather unlikely.

In this example, using traditional masses of the brain was useful because it is difficult to digitize landmarks within the endocranial cavity, thus making it difficult to obtain a good estimate of its centroid size. At the same time, brain masses are difficult to obtain, thus using a source of brain mass data from literature was helpful. When dealing with traditional biometric measurements such as weights (the same would apply to lengths or areas), size will logically be the dominant aspect of variation (Sokal and Rohlf 1995). This is important to take into account when using this kind of data with the 2B-PLS. In effect, this size variation is what the vector of the

first singular value is in fact singularizing (Fig. 7.3, see scatter-plots between body and brain mass against the SV of masses yielded by the 2B-PLS). For example, there is a nice linear relationship between both body mass and brain mass with the vector of masses from the first singular value. Thus, a parallel interpretation of the result is that cranial shape variation in birds is unlikely associated with size variation (i.e., it is possibly non allometric).

So, why is it that only some birds have a ventrally oriented foramen magnum as we humans do? To date even explaining this phenomenon in humans is still an intriguing issue. One of the most plausible ways that paleoanthropologists and morphologists are now using to explain why humans have a ventrally oriented foramen magnum is a structural rationale. Namely, researchers argue that the relationship between our enlarged brain volumes relative to our basicranial length is the motor of cranial change (see; Gould 1977). When the brain has grown as much as in humans it does so at the expense of the length of its bony support within the cranium, the cranial base in the midline. The phenomenon is thus seen as a spatial solution, since the tight packing of such an enlarged brain forces the cranial base to bend (Strait 1999; Ross et al. 2004; Lieberman et al. 2008). This bending of the base makes the nuchal plane to rotate ventrally, and therefore, the neck is forced to align vertical to the head instead of horizontal. The question of why do birds show this analogous feature in their cranium remains open, but it is plausible that encephalization in the above sense (the volume or mass of the brain relative to the extent of the base), instead of brain size alone (as results underscore), may be what causes the different neck attachments to the head in certain avian species. This rationale has never been addressed to explain cranial variation in birds, and is a new and suggestive path to continue investigating in avian evolutionary morphology.

Conclusions

The aim of this chapter has been to show the ability of the 2B-PLS to explore the interplay between shape and sources other than shape data, in particular, of mass data, but this could have been of lengths, areas or any other estimate of scale. The aim was to show how this approach can be helpful and fast to test complex hypotheses, and how it can be most helpful in settings which require an intricate experimental or data acquisition of one of the data sets. The example used has been the exploration of the potential association between brain size and cranial shape diversity between birds at high taxonomic scales of comparison. The 2B-PLS could not find any statistically significant covariation pattern between measurements of brain size and endocranial shape variation, suggesting that brain size alone is unlikely the sole factor causing the marked cranial diversity that we see between avian taxa.

References

Adams D, Rohlf FJ. 2000. Ecological character displacement in Plethodon: Biomechanical differences found from a geometric morphometric study. PNAS 97:4106–4111.

Adams DC, Slice DE, Rohlf FJ. 2004. Geometric morphometrics: ten years of progress following the "revolution". Ital J Zool 71:5–16.

Bookstein FL. 1991. Morphometric tools for landmark data: geometry and biology. 1st ed. New York: Cambridge University Press.

Bookstein FL, Gunz P, Mitteroecker P, Prossinger H, Schaefer K, Seidler H. 2003. Cranial integration in Homo: singular warps analysis of the midsagittal plane in ontogeny and evolution. J Hum Evol 44:167–187.

Duijm MJ. 1951. On the head posture in birds and its relation to some anatomical features. I–II Proc. Koninklijke Nederlandse Akademie van Wetenschappen, Ser C Biol Med 54:202–271.

Enlow DG, Hans MG. 1996. Essentials of facial growth. London: W.B. Sauders Co.

Gould SJ. 1977. Ontogeny and phylogeny. Cambridge, MA: Harvard University Press.

Gower JC. 1975. Generalized procrustes analysis. Psychometrika 40:33–51.

Harvey PH, Krebs JR. 1990. Comparing brains. Science 249:140–145.

Hesse E. 1907. Über der inneren knöcheren Bau des Volgelsechnabels. J Ornithol 55:185–248.

Iwaniuk AN, Nelson JE. 2002. Can endocranial volume be used brain size in birds? Can J Zool 80:16–23.

Lattin J, Carroll D, Green P. 2002. Analyzing multivariate data. Pacific Grove, CA: Duxbury Applied Sciences. Duxbury Press.

Lieberman DE, Hallgrímsson B, Liu W, Parsons TE, Jamniczky HA. 2008. Spatial packing, cranial base angulation, and craniofacial shape variation in the mammalian skull: testing a new model using mice. J Anat 212:720–735.

Livezey BC, Zusi RL. 2007. Higher-Order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. Biol J Linn Soc 149:1–139.

Marugán-Lobón J, Buscalioni AD. 2006. Avian skull morphological evolution: exploring exo- and endo-cranial covariation with two-block partial least squares. Zoology (Jena) 109: 217–230.

Marugán-Lobón, J, Buscalioni, AD. 2009. New insight on the anatomy and architecture of the avian neurocranium. Anat Rec 292(3):364–370.

Olson EC, Miller, RL. 1958. Morphological integration. Chicago: University of Chicago Press.

Pearson R. 1972. The avian brain. New York: Academic Press.

Portmann A. 1946. Études sur la cérébralisation chez les oiseaux 1. Alauda 14:2–20.

Portmann A. 1947. Études sur la cérébralisation chez les oiseaux 2. Alauda 15:1–15.

Rohlf FJ. 1990. Morphometrics. Ann Rev Ecol Syst 21:299–316.

Rohlf FJ, Corti M. 2000. The use of two-block partial least-squares to study covariation in shape. Syst Biol 49:740–753.

Rohlf FJ, Marcus LF. 1993. A revolution in morphometrics. Trends Ecol Evol 8:129–132.

Rohlf FJ, Slice DE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst Zool 39:40–59.

Ross CF, Henneberg M, Richard S, Ravosa MJ. 2004. Curvilinear and geometric modeling of basicranial flexion in a phylogenetic context: Is it adaptive? Is it constrained? J Hum Evol 46:185–213.

Roth VL, Mercer JM. 2000. Morphometrics in development and evolution. Amer Zool 40:801–810.

Sokal RR, Rohlf FJ. 1995. Biometry: the principles and practice of statistics in biological research. 3rd ed. New York: W. H. Freeman and Co.

Strait D. 1999. The scaling of basicranial flexion and length. J Hum Evol 37:701–719.

Striedter GF. 2005. Principles of brain evolution. MA: Sinauer Associates.

Windhager S, Slice DE, Shaefer K, Oberzaucher E, Thorstensen T, Grammer K. 2008. Face to face. The perception of automotive designs. Hum Nat 19(4):331–346.

Witmer LM, Chatterjee S, Franzosa J, Rowe T. 2003. Neuroanatomy of flying reptiles and implications for flight, posture and behaviour. Nature 425:950–953.

Zelditch ML, Swiderski D, Sheets D, Fink W. 2004. Geometric morphometrics for biologists: a primer. London: Elsevier.

Zusi R. 1993. Patterns of diversity in the avian skull. In: Hanken J, Hall B (Eds.), The Skull. Patterns of Structural and Systematic Diversity, vol. 2. Chicago: The University of Chicago Press, pp. 391–437.

Chapter 8
Biogeographic Analysis Using Geometric Morphometrics: Clines in Skull Size and Shape in a Widespread African Arboreal Monkey

Andrea Cardini, José Alexandre Felizola Diniz Filho, P. David Polly, and Sarah Elton

Idea and Aims

Despite the "renaissance" of biogeography in the last two decades with its central role in the study of biodiversity and evolution, and the "revolution" in morphometrics brought about by methods based on the analysis of Cartesian coordinates of anatomical landmarks, the use of geometric morphometrics in biogeographic studies has been rather limited. With this analysis we aim to provide an example of how geometric morphometrics can fruitfully be applied to the study of clinal variation in a widespread African monkey group by a simple extension of methods widely employed by macroecologists and biogeographers to multivariate shape data. Throughout the paper we aim to explain these techniques so that those who are new to them can use and adapt them for their own needs, in some cases providing specific instructions on how to perform certain operations in standard morphometrics and statistical software. Our hope is that this may stimulate morphometricians and scientists from other disciplines to explore geographic variation in size and shape using up-to-date geometric morphometric methods. The application of geometric morphometrics to ecological, biogeographic and phylogeographic studies has enormous potential for a thorough understanding of how form changes in space and time during evolution and in relation to genetic and environmental factors.

Introduction

"Biogeography ... asks a simple question: What lives where, and why?" (Parenti and Humphries 2004, pp. 899). This simple question is crucial to the understanding of how life evolves as "biogeographic patterns provide an organizing framework within which we may interpret biological data" (Parenti and Humphries 2004,

_____

A. Cardini (✉)
Dipartimento di Biologia Animale, Universitá di Modena e Reggio Emilia,
via Campi 213, 41100 Modena, Italy
e-mail: alcardini@interfree.it, cardini@unimo.it

pp. 899). Biogeography is thus central to the study of present and past life forms: it helps conservationists identify endemic groups and biodiversity hot-spots, provides essential knowledge for predicting how species distributions have changed in the past and may change in the future and gives important information for climatologists and policy makers. Together with the increasing availability of molecular data and advancements in statistical methods for spatial data analysis, these factors explain why biogeography has enjoyed a renaissance over the past two decades (Parenti and Humphries 2004). During the same two decades, morphometrics has been reinvigorated by the development of a new set of methods which preserve the geometry of shape in the study of biological forms and which take full advantage of new image technologies to obtain data and visualize results (Adams et al. 2004; Sanfilippo et al. in press). For this reason, geometric morphometrics, like biogeography, has been considered a "revolution" (Rohlf and Marcus 1993; Corti 1993).

Despite an increasing interest in using geometric morphometric (GMM) data in biogeographic analyses (Fadda and Corti 2001; Frost et al. 2003; Cardini et al. 2007, and references therein; Cardini and Elton 2009), few studies have tried to put together methods and ideas from the two fields. This is unfortunate becase both size and shape often vary with geography and are influenced by environmental factors, which themselves are related to geography. Only recently has effort been made to develop methods that illustrate multivariate variation in shape that are easily comprehendible within a geographic framework (Cardini and Elton 2009). To further pursue this aim, we describe in this chapter a detailed case study of variation in the African guenon monkey superspecies *Cercopithecus nictitans* (Grubb et al. 2003; Tosi et al. 2005), providing a step-by-step explanation of how to conduct biogeographic analyses of morphological data derived from GMM. Also, we suggest ways in which such analyses might be elaborated using more complex models, including those which take into account factors such as phylogenetic relatedness.

Evolutionary and ecological processes occur in explicit geographical contexts. In the late 1970s, several statistical techniques were introduced to investigate the mechanisms underlying spatial patterns in biological variables at different levels of the biological hierarchy (see Epperson 2003; Dormann et al. 2007, for recent reviews). Even more importantly, most of these techniques were developed to deal specifically with the issue of spatial autocorrelation. Spatial autocorrelation occurs when closer samples in geographical space tend to be more similar or dissimilar to each other for a given variable than expected by chance alone (Legendre and Legendre 1998). Spatial autocorrelation in biological variables has both endogenous and exogenous causes. An endogenous example is gene flow, which tends to spread through a population or among metapopulations in a wave-like fashion, causing higher genetic similarity, and thus morphological similarity, among neighboring locations than between distant ones (e.g., isolation-by-distance). Another possibility is that an exogenous factor causes the observed pattern, as is the case when a genetic variable responds to environmental variation (Fortin and Dale 2005; Kissling and Carl 2008).

When autocorrelation is viewed as an endogenous process, spatial autocorrelation analyses are usually applied to infer microevolutionary processes by exploring the spatial structure in the morphometric data. However in most cases the spatial patterns are also caused by exogenous factors, and it may thus be important to model the effects of the exogenous variables in order to statistically separate them from the endogenous ones. In this case, the main issue to be considered is that the existence of autocorrelation causes inferential statistical problems, since Type I errors in regression and correlation analyses are always inflated (see Legendre 1993; Diniz-Filho et al. 2003). More complicated models are thus often necessary to produce realistic inferences about patterns and processes.

Here we use African monkeys, the *Cercopithecus nictitans* group, to illustrate ways in which morphology can be investigated in a spatial framework. The *C. nictitans* group comprises all subspecies of the Greater spot-nosed monkey, *C. nictitans*, along with the closely related Blue monkey, *Cercopithecus mitis* (Kingdon 1988; Grubb et al. 2003). In nomenclatural terms, *C. nictitans* has priority over *C. mitis* (Grubb et al. 2003), but for the sake of clarity we hereinafter refer to the group as *C. mitis-nictitans*. The two groups are often considered to be a "superspecies" (Kingdon 1988; Grubb 2006), a clade comprising allopatric populations that are too morphologically distinct from each other to be included in a single species (Mayr 1963). The geographic ranges of *C. mitis* and *C. nictitans*, as expected for a superspecies, do not overlap. *Cercopithecus mitis* is the most geographically widespread of the arboreal guenon clade, distributed across sub-Saharan Africa as far north as Ethiopia, extending into eastern and southern South Africa by way of the Democratic Republic of Congo (DRC), Zambia and Mozambique (Kingdon et al. 2008). It is also found in Angola, in the far west of Africa, and Kenya and Tanzania in the far east of the continent (Kingdon et al. 2008). In common with other widespread African monkeys, it is polytypic and numerous subspecies are recognised (Kingdon et al. 2008). *Cercopithecus nictitans* is found in western Africa, including Nigeria, Cameroon, Congo and Gabon, with isolated populations in Liberia and Cote d'Ivoire, and Bioko Island (Oates and Groves 2008). The range of *C. nictitans* is nearly contiguous with *C. mitis* in northern DRC. The monophyly of the *C. mitis-nictitans* superspecies is supported by molecular data that indicate the two named groups are closely related (Tosi et al. 2005). Nonetheless, the taxonomy within the group is far from straightforward (Napier 1981). In molecular analysis, *C. mitis albogularis* groups with *C. nictitans* rather than with the other populations of *C. mitis* (Tosi et al. 2005), with Groves (2001) identifying *C. m. albogularis* as a distinct species.

The *C. mitis-nictitans* group is part of a Pliocene radiation of arboreal guenons (Tosi et al. 2005). *Cercopithecus mitis* exploits a wide variety of forest types across its extensive range, including thickets in areas of forest loss (Kingdon et al. 2008). *Cercopithecus mitis* is thus one of the most ecologically flexible arboreal guenons (Lawes 1990, 2002). *Cercopithecus nictitans* has a more restricted set of habitat preferences, being found in primary and secondary tropical forest (Oates and Groves 2008). Its fragmented distribution in the west of Africa

has been attributed to competition with *Cercopithecus diana* (Oates and Groves 2008). The arboreal guenons are likely to have speciated in the central African forest belt (Hamilton 1988), with *C. mitis* subsequently expanding its range into southern Africa, probably during the Pleistocene (Elton 2007). Despite its wide range, ecological flexibility and soft tissue variation, very little work has been undertaken on the hard tissue morphology of *C. mitis-nictitans*, especially in the context of geographic and environmental variation. Within the guenon clade as a whole, the allometry of skull shape appears largely to be conserved, although phylogenetic differences have been observed in some cases (Cardini and Elton 2008a, b). *Cercopithecus mitis* and *C. nictitans* are no exception to the conserved allometry, with their skull shapes being grouped with other medium to large-sized arboreal guenons (Cardini and Elton 2008a). Although on visual inspection many guenons seem to have a homogenous skull form, detailed GMM data allow for good species discrimination (Cardini and Elton 2008a). Nevertheless, when classification errors occurred, it was *C. mitis* specimens that were misclassified into *C. nictitans* with no reciprocal misclassification in the *C. nictitans* sample, though they were sometimes misclassified as other species (Cardini and Elton 2008a). Examining spatial and ecological variation may help to explain this pattern and shed further light on morphological differentiation within the *C. mitis-nictitans* superspecies.

In this study, we examine geographic variation in skull form of the *Cercopithecus mitis-nictitans* superspecies using linear and curvilinear regression models. Similar studies have been undertaken on the widely distributed African vervet (Cardini et al. 2007) and red colobus (Cardini and Elton 2009) monkeys. In both groups, a strong longitudinal, non-linear cline was identified in both size and shape. In vervets, morphological variation was also distributed along gradients of environmental variables. Precipitation emerged particularly strongly in these models, suggesting that habitat productivity affects size and shape, with larger animals living in higher rainfall and hence probably more productive environments. A similar finding has been reported for baboons (Dunbar 1990; Barrett and Henzi 1997b) and hartebeest (Capellini and Gosling 2007). In the red colobus, as with vervets, specimens in the far east of Africa were considerably smaller than elsewhere, although no study to date has specifically considered the role of environmental variables in influencing this pattern.

Our main aims in this paper are: (1) to investigate clinal variation in *C. mitis-nictitans*, taking into account possible non-linear patterns; (2) to summarize and visualize clines in size and shape within *C. mitis-nictitans*; (3) to examine environmental factors which may contribute to explain geographic variation in the superspecies; and (4) to partition size and shape into components (geographic, environmental, spatially structured environmental and residual) of ecogeographic variation. Throughout the paper we attempt to explain the methods so that those who are new to them can use and adapt them for their own needs, in some cases providing specific instructions on how to perform certain operations in standard morphometrics and statistical software.

Methods

Data Collection

The sample was derived from museum collections, comprising 122 specimens of *C. mitis* (56 females and 66 males) and 42 *C. nictitans* (22 females and 20 males), which had already been included in previous studies on guenons (Cardini and Elton 2008a, b, c). The specimens were all wild-caught and provenanced, with ninety-one localities represented. Data from the same locality were averaged and included in the analysis as means for those localities. A list of specimens with catalogue numbers is available upon request.

Three-dimensional coordinates of anatomical landmarks were directly collected by the same person on crania and mandibles using a 3D-digitizer (MicroScribe 3DX, Immersion Corporation). Landmarks were digitized only on the left side to avoid redundant information in symmetrical structures. The set (configuration) of 86 landmarks used for the analysis (Fig. 8.1) is described in Cardini et al. (2007) and several other papers from the same series of studies (Cardini and Elton 2007; Cardini and Elton 2008a, b, c). Landmarks on crania and mandibles were digitized separately, and aligned using the method described in Cardini et al. (2007). Measurement error and estimates of a small number of missing landmarks, described in Cardini et al. (2007) and Cardini and Elton (2008a), had negligible effects on the analysis.

Geometric Morphometrics

We used a geometric morphometric approach (Adams et al. 2004; Rohlf and Marcus 1993), an extensive introduction to which can be found in Zelditch et al. (2004). Geometric morphometric analyses were performed in the following computer programs: Morpheus (Slice 1999), NTSYSpc 2.2 V (Rohlf 2009), Morphologika (O'Higgins and Jones 2006). Statistical analyses were performed using SPSS 15.0 (2006) and NTSYSpc 2.2 V (Rohlf 2009).

A geometric morphometric analysis involves a series of main steps, described briefly here. The form of an organism (or its organs) is first captured by the Cartesian coordinates of a three-dimensional configuration of anatomical landmarks. Differences in landmark coordinates, due to the position of the specimens during the digitization process, are then removed, and size is standardized. This was achieved in our study by optimally superimposing landmark configurations using a process called generalized Procrustes analysis (GPA), which is based on a least-squares algorithm (Rohlf and Slice 1990). Centroid size (henceforth called "size" for brevity) is a measure of the dispersion of landmarks around their centroid and is computed as the square root of the sum of squared distances of all landmarks from the centroid. The new Cartesian coordinates obtained after the superimposition are the shape coordinates used for statistical comparisons of individuals. The shape differences between landmark configurations of two individuals can be summarized by
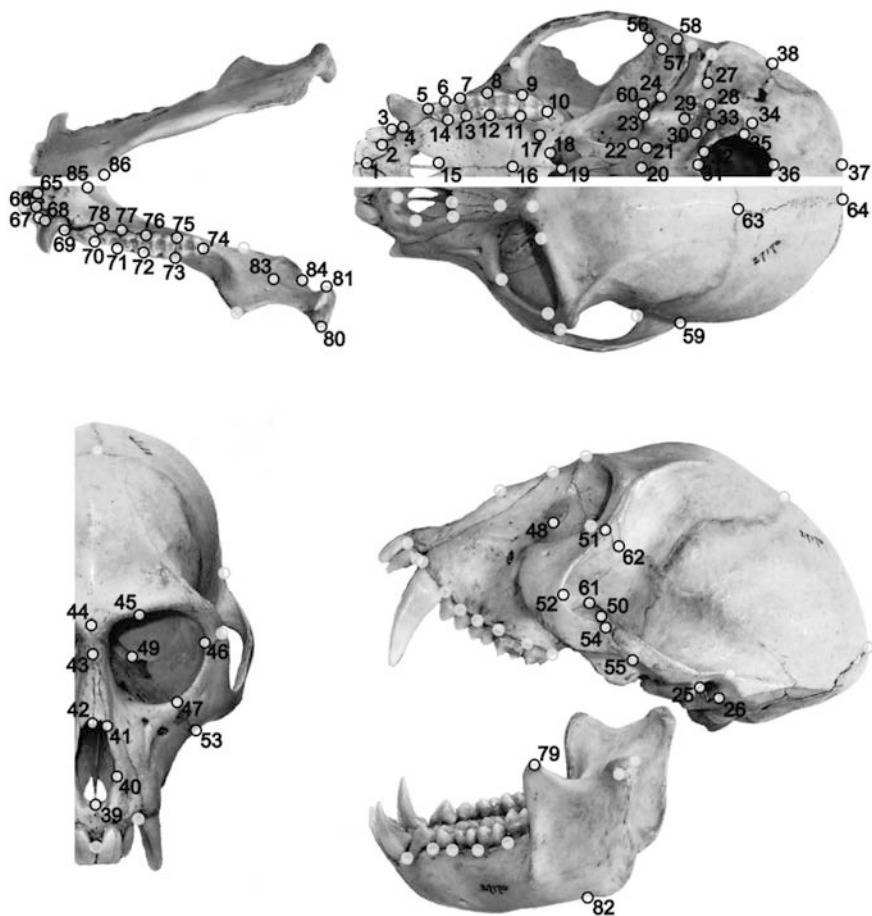
Fig. 8.1 Landmark configuration

their Procrustes distance, which is the square root of the sum of squared distances
between pairs of corresponding landmarks.

Correction for Sexual Dimorphism

Due to the strong sexual dimorphism in these monkeys and the need to maximise
sample size, a correction for sexual dimorphism was introduced. Within the sam-
ple for each species, the male-to-female differences between mean size and mean
shape were added to the size and shape of females respectively. In other words, we
standardized the female mean to equal the male mean and pooled the variance of
both sexes around the male mean (see also Cheverud 1995; Marroig and Cheverud

2004). Means were calculated in SPSS (size) and Morpheus (shape), and all other computations were done in a spreadsheet; size was restored by multiplying shapes by the corresponding size after the sex-correction. The simple approach we used to remove differences related to sex assumes that the pattern of sexual dimorphism is the same throughout the whole geographic range. Although previous studies on vervets (Cardini et al. 2007) and red colobus (Cardini and Elton 2009) suggested a generally good congruence between sexes in patterns of geographic variation of African primates, this assumption should be rigorously tested. This was not possible in our study sample because of the limited number of specimens and localities. Results must therefore be seen as preliminary and will need to be confirmed on a larger sample.

Mirroring one Side of the Skull, Removing Asymmetries, and Summarizing Shape

To visualise group differences more effectively using three-dimensional diagrams (see below), landmarks on the left side were reflected and tiny asymmetries on midplane landmarks, that have little relevance in a study concerning variation among individuals (Klingenberg et al. 2002), removed. This was achieved by performing the following operations:

(1) In a spreadsheet, we created a second set of configurations by inverting the sign of the axis that lay approximately perpendicular to the midplane.
(2) In Morpheus, we loaded all datasets (i.e., the original set of configurations together with the one created in the previous step), including all landmarks but superimposing configurations using midplane landmarks only (i.e., all landmarks are rescaled, translated and rotated but this is done in a way that optimally superimposes only midplane landmarks). A GPA on a subset of landmarks can be easily obtained in Morpheus by using the command "demote p 84" to exclude the 84th landmark, "demote p 83" to exclude the 83rd landmark and so on. It is advisable to first "demote" the last landmark to be excluded, then the penultimate, and so on. In this way, the numbering of landmarks is not altered and one can use the number in the original landmark configuration to tell the software which landmark to exclude. A batch file (ascii text file with extension "btc") can be easily written with the full list of landmarks to demote:

> demote p 84
> demote p 83
> demote p 82
> ….

(3) Data were then rescaled to restore size, and midplane landmarks averaged. Rescaling can be done either in Morpheus using the "super restore scale" command or in a spreadsheet by multiplying landmark coordinates from step 2 by the corresponding centroid size. Averages of midplane landmarks can be easily

obtained in a spreadsheet. Beware that when a subset of landmarks - like the 17 midplane landmarks in our study - is used for superimposition in Morpheus, these will become the first 17 landmarks in the output file.

(4) In the last step, the ordering of the landmarks was restored (this can be done in a spreadsheet by moving columns of x, y, z coordinates to the appropriate position) and the mirrored landmarks from the dataset created in the first step were appended to the set of 86 midplane and left-side landmarks. Thus, the final set comprised the original left-side landmarks, their mirror reflection and the averaged midplane landmarks.

The whole procedure removed small asymmetries in midplane landmarks and mirror reflected left side landmarks to produce a set of perfectly symmetric configurations. Redundancy in the "symmetrized" configuration was later removed by performing a principal component analysis of shape coordinates (see below).

Landmark shape coordinates are mathematically redundant because seven degrees of freedom are lost in a GPA of three dimensional data, plus they are often highly correlated (e.g., Adams et al. 2004; Rohlf and Marcus 1993). The number of variables used for the analysis of shape is standardly rreduced by including only the first principal components (PCs) of the shape coordinates. To do this, a principal components analysis (PCA) should be performed using the variance-covariance matrix, as coordinates have already been standardized, something that can be performed in Morphologika, NTSYSpc, SPSS or almost any other standard statistical software package. Here, the number of principal components to be analysed was selected by measuring the correlation between the matrix of Procrustes shape distances in the full shape space and pairwise Euclidean distances in the reduced shape space (5, 10, 15 principal components, and so on). Computations of distances and matrix correlations were done in NTSYSpc. By plotting correlation coefficients onto the number of components in a spreadsheet (Fig. 8.2) one can use this information in a way similar to scree plots to select how many variables summarize most shape variation (Fadda and Corti 2000; Cardini et al. 2007). The "elbow" in the plot suggests the minimum number of PCs to retain before the loss of information in the higher order PCs, which are excluded, is so large to appreciably change the relationships of specimens in the reduced shape space compared to the full Procrustes shape space. Thus, in our study, the first 20 principal components of shape explained 76.0% of total variance and had a correlation with distances in the full shape space of 0.986 and so were selected for use in all subsequent analyses.

(1) Analysis of Clines and Selection of the Best Model to Quantify Geographic Variation

Regressions of size and shape onto geographic coordinates were used to quantify clinal variation. This was done both using a simple regression onto latitude and longitude and by performing a trend surface analysis (TSA – Legendre and Legendre 1998; Ruggiero and Kitzberger 2004; Botes et al. 2006; Cardini et al. 2007; Cardini and Elton 2009). TSA is a curvilinear model that can take into account
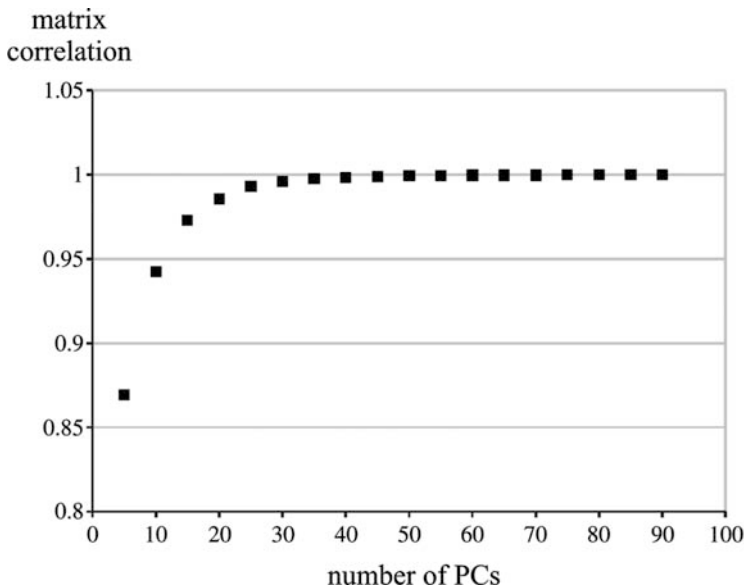
Fig. 8.2 Correlation between matrices of Euclidean distances computed from 5, 10, 20 etc. PCs and the matrix of Procrustes distances in the full shape space

non-linearities in the relationship between independent and dependent variables. Thus, size or shape variables were regressed in SPSS onto a third-order polynomial $(x, y, x^2, xy, y^2,$ and so on) of latitude and longitude, and non-significant terms were removed one by one starting with the largest and then repeating the regression until all terms of the multiple regression were significant. Although more sophisticated methods, such as eigenvector mapping (see Dormann et al. 2007; Griffith and Peres-Neto 2006), are available to represent geographic structure at multiple scales, the simple approach used here, based on broad-scale clines expressed by TSA, was effective in describing the clines and, at the same time, generally good at reducing residual autocorrelation (see below).

To select the best model, results of the TSA were compared with those of both simple linear models and the full third-order polynomial expansion of geographic coordinates. For this purpose, size or shape variables were regressed onto geographic coordinates, and the second order Akaike information criterion (AICc; Burnham and Anderson 1998; Mazerolle 2004) was used to compare the goodness of fit of the models. AICc is a measure based on information theory and derived from the concept of entropy in physics. Briefly, it measures the lack of fit of the data (sum of squared residuals in a regression) to a given model, where the model is penalized in proportion to the number of parameters it employs. Thus, compared to available alternatives, the best model is the one with the lowest AICc (AICcmin). The multivariate extension of AICc suggested by Burnham and Anderson (1998) is described in Cardini and Elton (2009). The relative level of support for different models was evaluated by $\Delta_{\mathrm{AICc}} = \mathrm{AICc} - \mathrm{AICc}_{\min}$ and Akaike weights (Burnham

and Anderson 1998). Akaike weights provide another measure of the strength of evidence (likelihood) for each model and approximately represent the ratio of Δ_{AICc} values for each model relative to the whole set of candidate models. Burnham and Anderson (1998) suggest that models with Δ_{AICc} values of 0–2 provide similar support, whereas $\Delta_{AICc} > 2$ indicate substantially less support than the best model.

All these analyses can be affected by residual autocorrelation, even when no significance tests are used (see Legendre 1993; Diniz-Filho et al. 2008). Residual autocorrelation can be tested using Moran's I coefficients (Legendre and Legendre 1998, pp. 714–721) in the univariate case (i.e., size) and multivariate Mantel correlograms (Legendre and Legendre 1998, pp. 736–738) in the multivariate case (i.e., shape). The tests estimate whether groups of localities that have approximately the same distance from one another (distance classes) are more similar than the average of all other localities based on the study variable(s). Ten distance classes were chosen so that there were an equal number of localities in each group and the geographic distances between pairs of localities in each group were similar. Distances between pairs of localities were smallest in the first group, greatest in the tenth group, and intermediate in the intermediate groups. In the absence of spatial autocorrelation, coefficients should not differ significantly from zero (i.e., morphological distances within a class are on average about the same as in any other class).

Moran's I coefficients for size were computed in SAM 3.0 (Rangel et al. 2006). Matrices of pairwise shape distances for Mantel tests were computed in NTSYSpc using SIMINT (option for Euclidean distances). Distance matrices from geographic coordinates were calculated in ArcView (Jenness 2005) and transformed into model matrices for each of the 10 geographic distance classes using SPSS ("recode into same variables" option with specific ranges for each distance class). Thus, all neighbouring localities within a given distance class were coded as 1 and the remainder values were set to zero. Mantel tests for the correlation between shape distances and model matrices were done in NTSYSpc. The sign of the matrix correlation r was inverted so that positive r implies positive autocorrelation. Significance of both Moran's I and matrix correlation was tested using 10,000 random permutation. A sequential Bonferroni correction using Holm's method (Howell 2002) was employed to control for the inflation of type I errors in multiple comparisons.

(2) Visualization of Clines in Size and Shape

Specimens were plotted according to geographic coordinates on a map of Africa using Arcview GIS 3.2 (1999). Clinal variation predicted by the selected model was illustrated with grey scale colour symbols on the map. Size, which is univariate, can be easily described by a single variable. Thus, grey symbols and contour plots of a tone proportional to the size of the skull predicted by geography were used. Contour plots based on clinal size help to visualize main trends in geographic variation in a way similar to altitudinal lines in terrain maps and can be easily computed in software for the analysis of geographic data like Arcview.

For the visualization of clines in shape, which is multivariate, we followed the method proposed by Cardini and Elton (2009), which summarises the main trends of

geographic variation and the corresponding shapes described by three dimensional anatomical landmarks using surface rendering. Thus:

(1) As for size, clinal shape variation predicted by the selected model was first computed and prediction scores saved. This can be done easily in NTSYSpc using the regression module but the SPSS general linear model for multivariate data could also be used, using predictors (i.e., geographic coordinates) as covariates.
(2) The variables describing the predicted cline in shape were then subjected to gsPCA ("geo-shape PCA", *sensu* Cardini and Elton 2009) in order to summarize most of the variation predicted by the model with a few variables. As before, the PCA is done using the variance covariance matrix.
(3) Eventually, variation along gsPCA axes (gsPC1 and gsPC2) was illustrated using:

  (a) grey scale colour symbols and contours of a tone proportional to the score of gsPC1 (or gsPC2) for individuals plotted on a map of Africa;
  (b) surface renderings of shapes corresponding to individuals at the opposite extremes of gsPC1 (or gsPC2). Surface rendering for main axes of clinal shape was obtained using Morphologika (O'Higgins and Jones 2006). Thus, the regression estimates of skull shape (the 20 PC scores) on geography were visualized as landmark configurations by adding to each mean coordinate the products of its eigenvectors and predicted PC scores. Predictions were computed in NTSYSpc (Rohlf 2009). The sum of predicted PC scores and mean coordinates was done in a spreadsheet and the resulting clinal shapes imported into Morphologika for visualization.

To aid interpretation of (b) relative to (a), shapes were shown using the same grey tone as for symbols and contours on the map. Thus, for instance, if lowest scores on gsPC1 were shown using light grey symbols on the map, light grey was also used for surface rendering of the shape predicted for the negative extreme of gsPC1. This allowed (a) mapping of clinal shape in a fashion similar to clinal size and (b) visualization of geographic shape variation as is commonly done in geometric morphometrics by using predictions for shapes along a vector (see Cardini and Elton 2009).

(3) Environmental Correlates

Size and shape were regressed onto environmental predictors and geographic coordinates. This was done in order to investigate which environmental variables may be related to changes in form once the effect of spatial distribution had been included in the model. The environmental variables used in the analysis were elevation, average annual temperature, rainfall and the Shannon rainfall diversity index, a measure of the differences in mean monthly rainfall over a 12 month period, with less seasonal environments represented by higher index values (Hill and Dunbar 2002). Elevation data were extracted from the SRTM 30 digital elevation model of Africa

(data available from USGS/EROS, Sioux Falls, SD) using the "extract values to points" procedure in ArcGIS 9.0 Spatial Analyst (2004). Climatic variables were taken from the Willmott and Matsuura database (see Cardini et al. 2007, for references) which provides data at 0.5 degree grids. As before, multiple (size) and multivariate multiple (shape) regressions can be done in NTSYSpc, SPSS and other commercial statistical software.

(4) Components of Ecomorphological Variance

Partial linear regression was used to assess the effects of spatial structuring of variables and estimate the amount of skull size or shape variation that could be attributed to different sets of factors (Legendre and Legendre 1998; Ruggiero and Kitzberger 2004; Botes et al. 2006). Thus, terms of geographic coordinates (spatial component) and environmental variables were combined and morphological variation was partitioned into four components: (1) non-environmental spatial (proportion of variance exclusively explained by geography); (2) spatially structured environmental (proportion of variance explained by both geography and environment); (3) non-spatial environmental (proportion of variance explained exclusively by environment); (4) unexplained variation (proportion of variance explained by the effect of other factors). Partial regression employs a simple additive model based on the amount of variance explained by different sets of variables and is closely related to path analysis (Caumul and Polly 2005; Crespi and Bookstein 1989). Thus, using the following abbreviations for components of variance

$$\mathbf{S} = (1), \mathbf{E} = (2), \mathbf{SE} = (3), \mathbf{res} = (4)$$

and referring to amounts of variances explained by the different regressions as

T_r = regression onto both spatial and environmental variables (i.e., total set of predictors);

S_r = regression onto spatial variables only (latitude and longitude or terms from the TSA);

E_r = regression onto environmental variables only (elevation, temperature and so on)

we have

$$T_r = \mathbf{S} + \mathbf{E} + \mathbf{SE}$$
$$S_r = \mathbf{S} + \mathbf{SE}$$
$$E_r = \mathbf{E} + \mathbf{SE}$$

By making the appropriate substitutions in the equations, one gets estimates of all components:

$$T_r = S + E_r \text{ hence } S = T_r - E_r$$
$$T_r = E + S_r \text{ hence } E = T_r - S_r$$

which makes

$$SE = T_r - (S + E)$$

and finally

$$res = 100 - T_r$$

(variances are here expressed as percentages)

Jackknife confidence intervals (Manly 1997) were also computed, in NTSYSpc, to assess the reliability of estimates of explained variance. The jackknife requires the iterative removal of each specimen from its sample before the analysis is repeated and the parameter estimated. Thus, for the full regression model (T_r) for instance, the 91st specimen was removed, the regression done on the subsample of 90 specimens and the corresponding percentage of variance explained by the computed predictors. This was repeated after removing the 90th specimen, then the 89th and so on until, one at a time, all specimens were excluded from the regression. Finally, the standard deviation of explained variances in the 91 regressions computed as described above was used to estimate the standard error (SE) of T_r. Assuming normality in the distribution of explained variances from the jackknife subsamples, 99% confidence intervals are computed as $T_r \pm 2.575$ x SE(T_r). The same procedure was used to estimate confidence intervals of S_r and E_r, and all other terms of the partial regression. Jackknives tend to underestimate SE (Manly 1997). To mitigate this problem we chose 99% confidence intervals instead of 95%. The jackknife was used instead of bootstrapping (random sampling with replacement) to avoid having multiple data points for the same locality, which would have been inconsistent with our approach of averaging size and shape of specimens with the same provenance. Percentages and confidence intervals were computed in a spreadsheet after jackknife results were imported.

Results

1–2. Models of Clinal Variation, and Visualization of Clines in Size and Shape

Results of the regressions of size and shape onto geographic predictors are shown in Table 8.1. Clinal variation was highly significant for both size and shape. For size, the model with the highest likelihood was derived from the TSA and effectively removed autocorrelation (-0.110 < Moran's I < 0.130, non-significant after a sequential Bonferroni correction). Other models had $\Delta_{AICc} > 8$. TSA explained about 50% of variation in size. The non-linear trend showed a strong longitudinal

Table 8.1 Clinal variation in size and shape: comparison of different models[a] using percentages of shape variance explained (% ex.), significance in F tests, second order Akaike information criterion (AICc), Delta AICc (Δ_{AICc} = AICc– $AICc_{min}$) and Akaike weights (W_i)

Form	Predictors	% ex.	Wilks'λ	F	df_1	df_2	P	AICc	Δ_{AICc}	W_i
Size	x^2, x^3, xy, y	50.8	–	20.170	4	86	9.6×10^{-12}	240.7	0.0	0.987
	full polynomial	70.8	–	9.044	9	81	2.5×10^{-9}	249.4	8.7	0.013
	x, y	15.7	–	8.177	2	88	0.001	256.1	15.4	0.000
Shape	x^2, x^3, xy, y	16.6	0.038	4.305	80	266.7	1.1×10^{-16}	–225.0	11.7	0.003
	full polynomial	19.8	0.013	2.041	180	529.7	3.4×10^{-10}	–224.0	12.7	0.002
	x, y	9.2	0.221	3.870	40	138.0	1.8×10^{-9}	–236.7	0.0	0.995

[a]In this and the next table, geographic predictors are shown using x, for longitude, and y, for latitude.

component. However, a north-to-south gradient in size was also evident. Figure 8.3 summarizes clinal size. Largest skulls were found in the central-south region of the *C. mitis-nictitans* geographic range. Size becomes progressively smaller to the north-east and to the north-west. Smallest individuals were found in Liberia and Sierra Leone. East of the African Rift, the gradient in size reduction was particularly steep.

For shape, the model with the highest likelihood was the linear regression onto latitude and longitude. Curvilinear models had $\Delta_{AICc} > 11$. Thus, about 9% of variation in shape was linearly related to geography although the regression was not completely effective in removing residual autocorrelation, as indicated by significance of two out of 10 distance classes after a sequential Bonferroni correction (distance class 678–857 km, $r = -0.0863$, $p = 0.0011$; distance class 2384–3221 km, $r < 0.1294$, $P = 0.0025$). The main component (~80% of clinal shape) of clinal shape was longitudinal. This component corresponds to gsPC1 and is illustrated in Fig. 8.4. To the west faces tend to be broad and short, brow ridges are relatively inconspicuous, neurocrania somewhat round in shape and nasals short. To the east, the trend is reverted. Faces are long with a narrow snout, brow ridges are prominent and zygomatic arches expanded, the temporal fossa is deep; also, the cranial vault is oval-shaped, and the mandible has a massive appearance compared to those from the west side of the range. The latitudinal component (gsPC2; Fig. 8.5) was modest and accounted for the remaining 20% of clinal shape variation. This component seemed mostly related to a lateral compression of the skull and a propensity to having longer faces in the south than in the north.

3–4. Environmental Correlates and Ecomorphological Components of Size and Shape Variation

When environmental predictors were included in the analysis together with geographic coordinates from the previous analyses, none were significant for either size
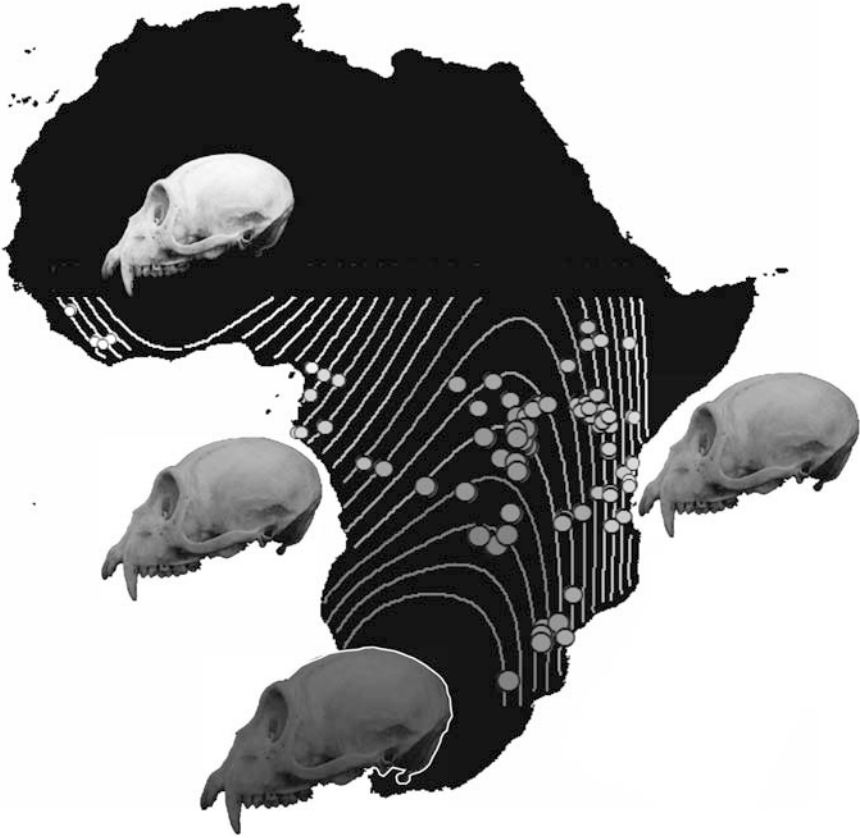
Fig. 8.3 Clinal size. Grey scale of contour lines and symbols, plus size of symbols, are according to increasing skull size. Largest skulls are found in south and central tropical Africa, smallest ones are in the westernmost part of the superspecies range (Liberia, Sierra Leone), medium sized skulls are both in the west (e.g., Cameroon and Gabon) and the east (e.g., Kenya, Tanzania and Ethiopia) of central tropical Africa. The main trend in size is summarized by pictures of crania (grey scale and size approximately proportional to skull size; pictures are from http://1kai.dokkyomed.ac.jp/mammal/en/mammal.html)

or shape (Table 8.2). Temperature, however, was marginally significant for shape. Bivariate correlations also suggested that for size temperature ($r = -0.427$, $P < 0.001$) and also elevation ($r = 0.253$, $P < 0.05$) may have some effect, but this was too small to be significant in the full model.

Regressions used to compute percentages of variance explained by different factors are shown in Table 8.3. The increase (from about 48 to 51%) in the percentage of variance explained by the model when environmental predictors are added to geographic coordinates was very modest. That the environment alone was a poor predictor was emphasized by the results of the partial regression analysis (Table 8.4). The percentage of variance explained exclusively by environmental variables was
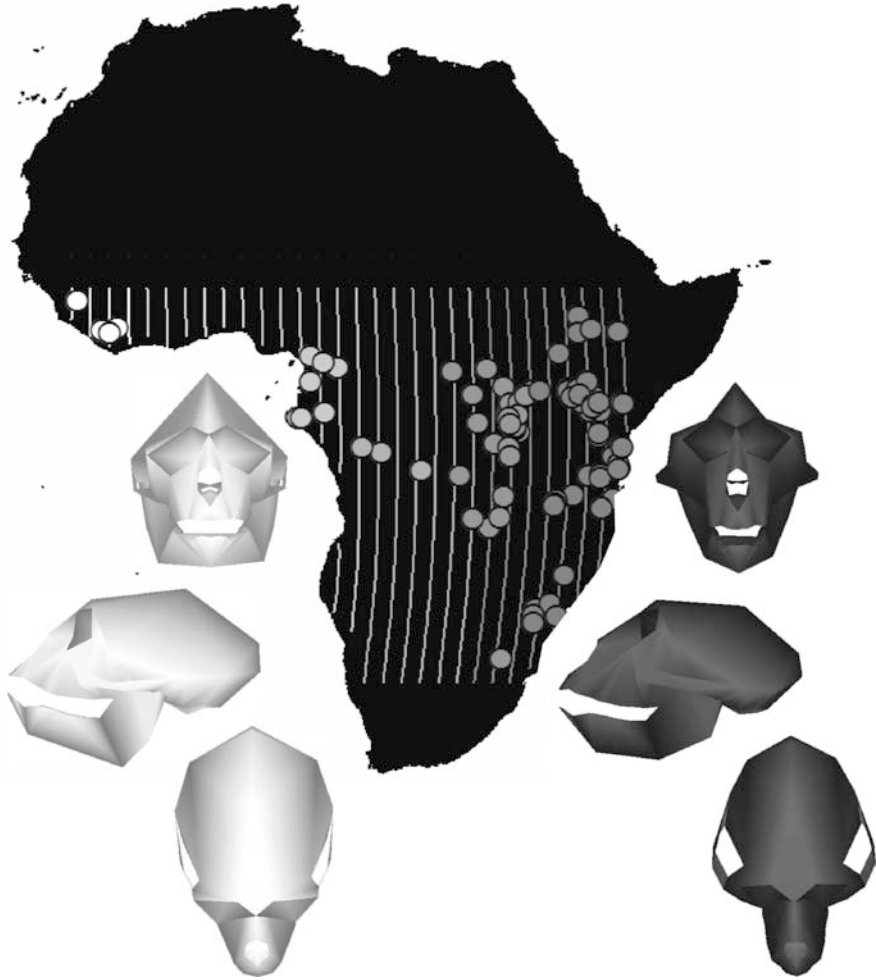
Fig. 8.4 Clinal shape. The first major axis (gsPC1) is shown which summarizes 80.3% of clinal variation in shape and has a clear longitudinal direction. Grey scale of contours and symbols on the map are proportional to gsPC scores. Shape changes at extremes of the axis are shown with surface renderings using three views (frontal, lateral and dorsal). Consistently with the grey scale of symbols on the map, dark and light grey tones are used for shapes corresponding to the opposite extremes of gsPC1. The same convention for symbols, contours and shapes is used in the next figure for the second major axis of clinal shape (Fig. 8.5)

tiny compared to percentages of variance related to the spatially structured environmental and purely spatial components (8–12 times larger respectively). Thus, size showed a pattern which was partly related to environmental variation along a geographic gradient and partly dependent on geographic distance regardless of the environment. In a different pattern to that evident for size, when environmental
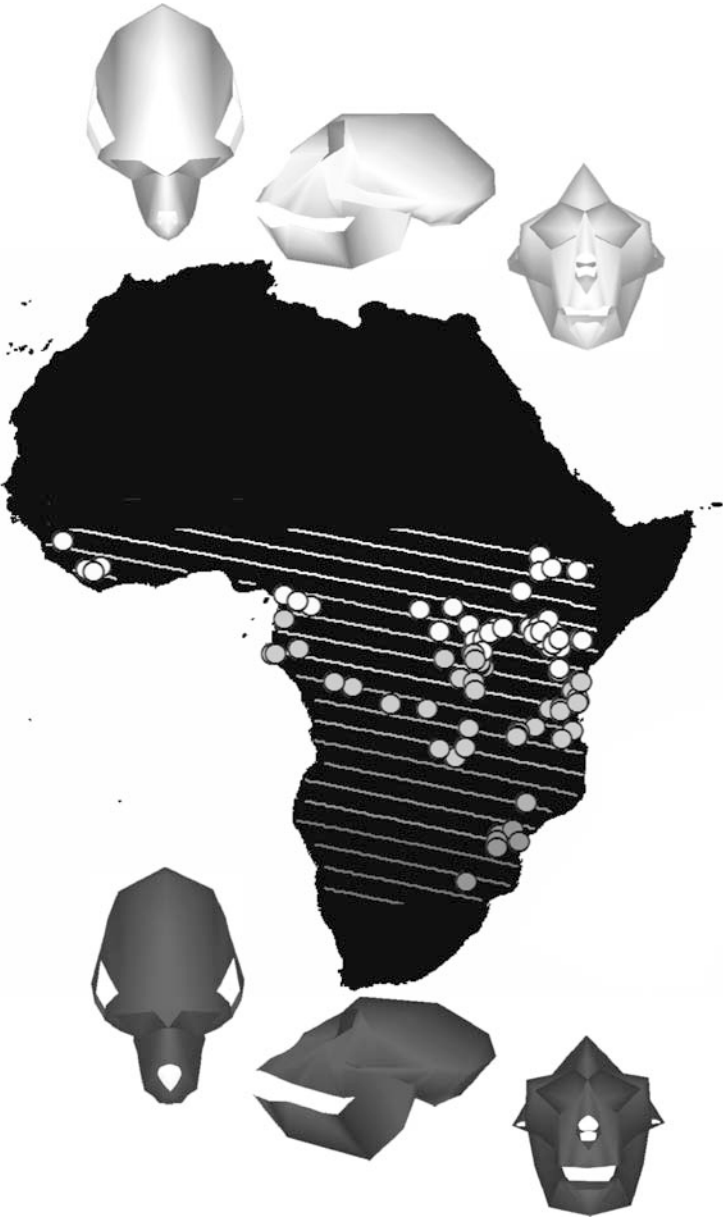
Fig. 8.5 Clinal shape. The second major axis (gsPC2) is shown which summarizes 19.7% of clinal variation in shape and is mostly according to a latitudinal gradient

Table 8.2 Regressions on environmental correlates and best fit geographic predictors

Form	Predictors	Wilks' λ	F	df$_1$	df$_2$	P
Size	Intercept	–	71.844	1	82	7.9×10^{-13}
	x^2	–	13.737	1	82	3.8×10^{-4}
	x^3	–	16.772	1	82	9.8×10^{-5}
	xy	–	3.225	1	82	0.076
	y	–	6.383	1	82	0.013
	precipitation	–	0.138	1	82	0.711
	temperature	–	1.639	1	82	0.204
	Shannon index	–	0.000	1	82	0.996
	elevation	–	0.152	1	82	0.698
Shape	intercept	0.868	0.494	20	65	0.960
	x	0.528	2.910	20	65	0.001
	y	0.716	1.292	20	65	0.217
	precipitation	0.757	1.044	20	65	0.428
	temperature	0.656	1.704	20	65	0.056
	Shannon index	0.849	0.579	20	65	0.913
	elevation	0.797	0.828	20	65	0.672

variables together with latitude and longitude were used to predict variation in shape (Table 8.3), the amount of variance explained increased 1.5 times. In the partial regression, about the same percentage of variance (~5%) was explained by geography alone, the environment alone or both of them congruently.

Differences in patterns of size and shape variation relative to geography and environmental predictors are more easily appreciated using profile plots (Figure 8.6) for components of variances. The dominance of geography and the spatially structured environmental variation was evident for size. In contrast, confidence intervals of the three components largely overlapped for shape.

Table 8.3 Regressions used to compute partial regressions (see next Table)

Form	Model	Variables	Test Wilks' λ	F	df$_1$	df$_2$	P	% Explained[a] obs.	lower	Upper
Size	Full	geography + environment	–	10.569	8	82	4.9×10^{-10}	50.8	48.7	52.9
	Geography	x^2, x^3, xy, y	–	20.170	4	86	9.6×10^{-12}	48.4	46.4	50.4
	environment	all 4 environmental	–	5.860	4	86	3.2×10^{-4}	21.4	18.9	23.9
Shape	Full	geography + environment	0.053	2.109	120	383.1	4.2×10^{-8}	14.1	13.4	14.9
	Geography	x, y	0.221	3.870	40	138.0	1.8×10^{-9}	9.2	8.6	9.9
	environment	all 4 environmental	0.137	2.178	80	266.7	1.9×10^{-6}	10.3	8.9	11.7

[a]99% jackknife confidence intervals.

Table 8.4 Partial regressions of size and shape onto geography and environment: percentages of variance explained by different components

		% Explained[a]		
Form	Variation component	obs.	lower	upper
Size	Environment	2.4	1.7	3.1
	common	19.0	16.9	21.1
	geography	29.4	26.7	32.1
	unexplained	49.2	47.1	51.3
Shape	Environment	4.9	4.5	5.2
	common	5.4	3.9	6.9
	geography	3.9	2.1	5.6
	unexplained	85.9	85.1	86.6

[a]99% jackknife confidence intervals.
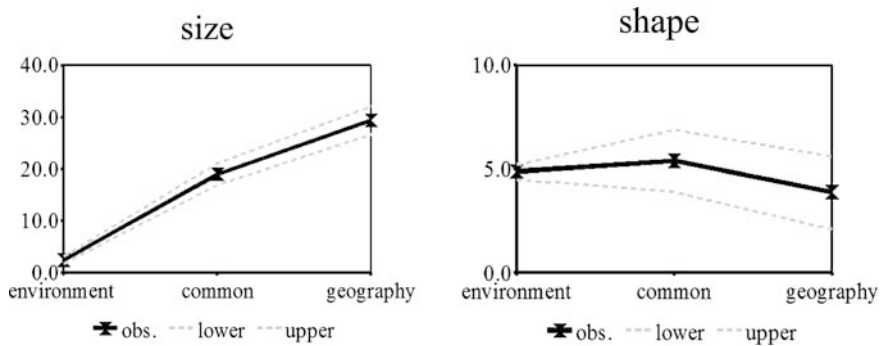
Fig. 8.6 Partial regressions of size and shape onto geography and environment: percentages of variance explained by different components and 99% jackknife confidence intervals illustrated using profile plots

Discussion

The strong spatial component to skull morphology in *C. mitis-nictitans* reinforces the emerging trend that geography is an important determinant of size and shape within widespread African primates, both arboreal and terrestrial (Frost et al. 2003; Cardini et al. 2007; Nowak et al. 2008; Cardini and Elton 2009). The observed size cline is non-linear, with a clear longitudinal component. There is an especially marked size reduction in the far east of the *C. mitis-nictitans* range, with another reduction to the west. Latitudinally, there is also a size gradient, although this is less steep, with individuals to the north of the range being only slightly smaller than those in the south. Essentially, the largest individuals - those at the "top of the hill" on the gradient plots - are found in the centre of the distribution, in the central African forests.

The *C. mitis-nictitans* superspecies is polytypic (Grubb et al. 2003), and marked intraspecific variation has been observed in its dentition (Vitzthum 1990) and pelage (Groves 2001). Our data indicate that skull shapes vary across the range, with specimens from the west having rounded neurocrania and short broad faces, with the trend reversed to the east. Southerly specimens have longer faces than those in the north. Allometry probably contributes to differences in morphology, with smaller individuals showing more paedomorphic features. This has been observed in several mammal species and is discussed in detail elsewhere (Cardini and Elton 2009). The obvious shape trend is longitudinal but it is likely that shape does not vary in a simple linear way, hinted at by the pattern seen on gsPC2. Larger, well provenced samples taken from the whole *C. mitis-nictitans* distribution are necessary to investigate this in detail. It is thus worth noting that when examining spatial trends in shape, which is multivariate, large samples are necessary to identify detailed variations in morphology in taxa with a large geographic range.

The *Cercopithecus mitis-nictitans* group is one of several widely distributed Old World monkey taxa found in Africa. Others include vervet monkeys (*Chlorocebus aethiops* spp.), common baboons (*Papio hamadryas* spp.), black and white colobus (*Colobus* sp.) and red colobus (*Piliocolobus* sp.). Of these, similar studies to the one reported here have been conducted on vervets (Cardini et al. 2007) and red colobus (Cardini and Elton 2009). Some interesting, if tentative, patterns emerge from the comparison of skull variation between these two groups and *C. mitis-nictitans*. All show a marked decrease in skull size in specimens east of the African Rift Valley lakes. In the rest of Africa, clinal variation in size is especially similar in *Piliocolobus* and *C. mitis-nictitans*, with specimens from central Africa being remarkably larger than those from west Africa. This contrasts with the trend in *Chlorocebus aethiops*, in which the largest forms are found in West Africa (Cardini et al. 2007).

It is likely that a complex, interacting array of factors influence morphological variation in African primates. These may contribute to size and shape differentiation through adaptation as well as developmental plasticity. Influential factors include (but are not confined to) phylogenetic history, environmental variables such as temperature and rainfall, insularity and other small population effects, stochastic processes and diet. In sympatric but distantly-related primates, shared behaviours and ecologies may contribute to parallel morphological evolution. Although *C. mitis-nictitans* and *Piliocolobus* do not share a particularly close evolutionary relationship, with molecular data indicating that colobines and cercopithecines diverged around 16 Ma (Raaum et al. 2005), both taxa are arboreal, found in similar regions and have a Pliocene origin (Tosi et al. 2005; Ting 2008). It is therefore probable that they have been subject to similar environmental pressures and hence parallelism.

During the Pleistocene climatic fluctuations, forest cover in tropical Africa would have fluctuated, with vicariance events promoting speciation within African forest primates (Hamilton 1988). Forest fragmentation, alongside competition with other arboreal mammals, including primates, has also very probably influenced distributions and population histories. Kingdon (1990) has equated African forest fragments

to islands, and it is possible that insularity has impacted upon the morphological evolution of both *C. mitis-nictitans* and red colobus (Nowak et al. 2008; Cardini and Elton 2009). Extremely small *C. nictitans* specimens are found in the far west of their distribution, in populations geographically isolated from the rest of the sub-species range (Oates and Groves 2008). It cannot be discounted that their small size is causally related to their geographic and possibly reproductive isolation. The fragmentation of the *C. nictitans* range may also help to explain a previous observation (Cardini and Elton 2008a) that their skulls are more morphologically heterogeneous that those of *C. mitis*. Similarly, *Piliocolobus* has a discontinuous distribution within West Africa (Oates et al. 2008), which again may have affected size (Cardini and Elton 2009). One possible mechanistic explanation for decreased size in isolated populations is that smaller size reduces competition in resource-limited environments (Lomolino 2005). Arboreal primates lack the ability to move easily between widely separated forest patches, so habitat fragmentation is more likely to create isolated populations and reduce gene flow. Terrestrial or semi ter-restrial primates, like the vervets, have fewer constraints on movement, and may be able to avoid strong competition within restricted areas (Cardini and Elton 2009). This may explain why vervets diverge from the trend identified in red colobus and *C. mitis-nictitans*.

Differences in the results of regression analyses indicate that although there is a strong spatial component to skull variation in *C. mitis-nictitans*, *Piliocolobus* and *Chlorocebus aethiops*, environmental influences may vary. In *C. mitis-nictitans*, precipitation and annual rainfall diversity (as indicated by the Shannon index) are not significant predictors for size or shape, although temperature is on the bound-aries of significance as a predictor for shape. In vervets, skull shape is affected by average annual rainfall, but to less of an extent than size, probably because size responds more quickly to environmental variation than shape, which is inherently more complex (Cardini et al. 2007). Small vervet size may be related to drier and less productive habitats (Cardini et al. 2007). Similar suggestions have been made for baboons (Dunbar 1990; Barrett and Henzi 1997a) and hartebeest (Capellini and Gosling 2007). Within vervets, however, the regression models for size dif-fer slightly in males and females, with rainfall having more of an effect on male size and temperature contributing to female size. It is therefore possible that the use of a corrected, pooled sex sample for *C. mitis-nictitans* may conceal subtle differ-ences in morphology that are related to ecological variations between males and females and may be detectable in larger samples. Indeed, the partial regressions are fairly similar in the two taxa, with the common component of geography and envi-ronment accounting for a much higher percentage of variance than environment on its own (Cardini et al. 2007). The relationhips between environmental, geographic and morphological variables are yet to be specifically tested in red colobus. Doing so and comparing with the work described here may help to elucidate the patterns observed in the skull morphology of arboreal and terrestrial African primates with large geographic ranges.

Our approach to spatial analysis of morphology aims to distinguish among differ-ent factors that contribute to geographic variation, including the exogenous effects

of environment and the endogenous effects of population structure and gene flow, with the special requirement of being able to visualize the components of multivariate geometric shape that are associated with different factors. Ours is not the only approach to this problem, but it is closely related and, we argue, more sophisticated than most. An approach that also partitioned morphological variance among exogenous and endogenous factors in a geographic setting using path analysis was Caumul and Polly's (2005) partitioning of variation among climate, diet, vegetation, size and phylogenetic factors. While the partitioning of variance approach of these authors was similar to what we describe here and could have been used to model shape geographically, geographic analysis was neither the aim of that study nor did it take advantage of the geographically appropriate curvilinear TSA models. Two-block partial least squares (2B-PLS) has also been used to analyze the effects of geography and climate on geometric morphometric measurements of morphology (e.g. Fadda and Corti 2001; Rychlik et al. 2006) but with similar differences to our approach.

Phylogenetic relatedness is a factor of interest to many biologists and one which may contribute importantly to patterns of geographic variation. Phylogeny is an example of an endogenous factor that grows out of the spatial autocorrelation introduced by population structure and gene flow. Closely related species are expected to be more similar to one another, regardless of where they live, than are distantly related species. At the time of speciation, the two ancestral populations of an allopatrically probably live in similar environments and have similar morphologies, but are geographically separated (Peterson et al. 1999). At the incipient stage of speciation, perhaps similar to the current stage of the *Cercopithecus mitis-nictitans* superspecies, the effects of phylogeny may not be much different from the effects of population structure and gene flow. However, in situations where distinct species with long geological histories are studied, phylogenetic effects may be more important. Over time, the ranges of distinct species will change as the result of migrations, environmental changes, and chance. Two related species that were once geographically close together may become separated by long distances into different environments, yet they will carry with them a common component of morphological variance that is due to their shared ancestry; as each species adapts to its new environment, the endogenous phylogenetic component will shrink in favour of the exogeneous environmental one. The phylogenetic component of variance is thus a dynamic one that interacts with the other factors that we considered in this study (e.g., Polly 2008). One way to deal with phylogenetic autocorrelation in morphology is to remove it by subtracting the variance that can be associated with phylogenetic relationships in much the same way that we substracted the variation due to sexual dimorphism. In other words, one could regress out the morphological variation due to phylogeny and examine only the residual variation using the methods described by Martins and Hansen (1997). This can be done in the program COMPARE, which has a web-based JAVA interface (Martins 2004). PC scores can be used as variables in the "taxon means" part of the program. The PGLS-Ancestor module will calculate ancestral values for each PC axis at every node on a tree. The ancestral value of taxa can be subtracted from their species means to standardize for the effects of

phylogeny in exactly the same way as we standardized for the effects of sex. This approach is not without problems, however, because often the geographic location of species and the climates they are adapted to are also phylogenetically correlated, which means that some of the geographic effects that one might want to study are being removed prematurely. A better approach would be to include phylogeny as a fifth regression term in the partial correlation analysis we presented above. To do this, one would add the phylogenetic weighting models presented by Martins and Hansen (1997) to the partial regression analysis, allowing phylogenetic factors to compete equally in the same partitioning of variance as the geographic and environmental effects.

Shape data are often analysed using simple models because their multivariate nature makes it more difficult to do complex analyses and because many statistical analyses are not yet available in user-friendly geometric morphometric programs. Besides statistical testing, results from the analysis of shape need to be effectively visualized, and this adds a futher layer of computational complexity. Our study shows how biogeographic studies using methods which allow consideration of possible non-linearities can be easily applied to size and shape data by using a combination of some of the most widely used commercial statistical packages together with freeware programs for geometric morphometric analyses.

The application of geometric morphometrics in biogeography is still in its infancy and much more need to be done to take full advantage of methods developed in this discipline over the last two decades. For instance, spatial autocorrelation may be harder to control for using spatial filters when the number of variables is large. This may help to explain why the best fit regression model effectively removed autocorrelation from size but did not do it as well for shape. Most of the biogeographic literature, however, either focuses on single dependent variables (i.e., univariate or univariate multiple analyses) or lacks explict tests of autocorrelation in multivariate data, thus limiting the usefulness of previous studies in the interpretation of our results.

The main aim of this study was to provide a step-by-step example of biogeographic analysis using geometric morphometrics which may stimulate morphometricians and scientists from other disciplines to investigate geographic variation in size and shape using up-to-date methods for the analysis of form using Cartesian coordinates of anatomical landmarks. Although more studies on how best to control and test for autocorrelation due to spatial or phylogenetic structure in multivariate shape data are needed, the application of geometric morphometrics to ecological, biogeographic and phylogeographic studies has an enormous potential for a thorough understanding of how form changes in space and time during evolution and in relation to genetic and environmental factors.

References

Adams DC, Rohlf FJ, Slice DE (2004) Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology 71: 5–16.

ArcGIS 9.0 (2004) Environmental Research Institute Inc., Redlands, CA.

ArcView GIS 3.2 (1999) Environmental Research Institute Inc., Redlands, CA.

Barrett L, Henzi SP (1997a) An interpopulation comparison of body weight in chacma baboons. South African Journal of Science 93: 436–438. Biogeography 32:1683–1699.

Barrett L, Henzi SP (1997b) Environmental determinants of body weight in chacma baboons. South African Journal of Science 93: 436–438.

Botes A, MA, McGeoch H, Robertson G, van Niekerk A, Davids HP Chown SL (2006) Ants, altitude and change in the northern Cape Floristic Region. Journal of Biogeography 33: 71–90.

Burnham KP, Anderson DR (1998) Model selection and inference: a practical information-theoretic approach. New York: Springer.

Capellini I, Gosling LM (2007) Habitat primary production and the evolution of body size within the hartebeest clade. Biological Journal of the Linnean Society 92: 431–440.

Cardini A, Elton S (2007) Sample size and sampling error in geometric morphometric studies of size and shape. Zoomorphology 126: 121–134.

Cardini A, Elton S (2008a) Variation in guenon skulls I: species divergence, ecological and genetic differences. Journal of Human Evolution 54: 615–637.

Cardini A, Elton S (2008b) Variation in guenon skulls II: sexual dimorphism. Journal of Human Evolution 54: 638–647.

Cardini A, Elton S (2008c) Does the skull carry a phylogenetic signal? Evolution and modularity and in the guenons. Biological Journal of the Linnean Society 93: 813–834.

Cardini A, Elton S (2009) Geographic and taxonomic influences on cranial variation in red colobus monkeys (Primates, Colobinae): introducing a new approach to 'morph' monkeys. Global Ecology and Biogeography DOI: 10.1111/j.1466-8238.2008.00432.x.

Cardini A, Jansson A-U, Elton S (2007) Ecomorphology of vervet monkeys: a geometric morpho-metric approach to the study of clinal variation. Journal of Biogeography 34: 1663–1678.

Caumul R, Polly PD (2005) Phylogenetic and environmental components of morphological vari-ation: skull, mandible and molar shape in marmots (*Marmota*, Rodentia). Evolution 59: 2460–2472.

Cheverud JM (1995) Morphological integration in the saddle-back tamarin (Saguinus fuscicollis) cranium. American Naturalist 145: 63–89.

Corti M (1993) Geometric morphometrics: An extension to the revolution. Trends in Ecology and Evolution 8(2): 302–303.

Crespi BJ, Bookstein FL (1989) A path-analytic model for the measurement of selection on morphology. Evolution 43:18–28.

Diniz-Filho JAF, Bini LM, Hawkins BA (2003) Spatial autocorrelation and red herrings in geographical ecology. Global Ecol Biogeogr 12: 53–64.

Diniz-Filho JAF, Rangel TFLVB and Bini LM (2008) Model selection and information theory in geographical ecology. Global Ecol Biogeogr 17: 479–488.

Dormann CF, McPherson J, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto P, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of distributional species data: a review. Ecography 30: 609–628.

Dunbar RIM (1990) Environmental determinants of intraspecific variation in body weight in baboons (Papio spp). Journal of Zoology, London 220: 157–169.

Elton S (2007) Environmental correlates of the cercopithecoid radiations. Folia Primatologica 78: 344–364.

Epperson BK (2003) Geographical genetics. Princeton University press, Princeton, 376 pp.

Fadda C, Corti M (2000) Three dimensional geometric morphometric study of the Ethiopian Myomys – Stenocephalemys complex (murinae, rodentta). Hystrix 10(2): 131–143.

Fadda C, Corti M (2001) Three-dimensional geometric morphometrics of Arvicanthis: implications for systematics and taxonomy. Journal of Zoological Systematics and Evolutionary Research 39: 235–245.

Fortin M-J, Dale MRT (2005) Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, 382 pp.

Frost SR, Marcus LF, Bookstein FL, Reddy DP, Delson E (2003) Cranial allometry, phylogeography, and systematics of large-bodied papionins (Primates: Cercopithecinae) inferred from geometric morphometric analysis of landmark data. The Anatomical Record 275A: 1048–1072.

Griffith DA, Peres-Neto P (2006) Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. Ecology 87: 2603–2613.

Groves CP (2001) Primate taxonomy. Smithsonian Institution Press, Washington, DC.

Grubb P (2006) Geospecies and superspecies in the african primate fauna. Primate Conservation 20: 75–78

Grubb P, Butynski TM, Oates JF, Bearder SK, Disotell TR, Groves CP, Struhsaker TT (2003) Assessment of the diversity of African primates. International Journal of Primatology 24: 1301–1357.

Hamilton AC (1988) Guenon evolution and forest history. In: A primate radiation: Evolutionary biology of the African guenons (Gaultier-Hion A, Bourliere F, Gautier J-P, Kingdon J, eds.), Cambridge, Cambridge University Press, 13–34 pp.

Hill RA, Dunbar RIM (2002) Climatic determinants of diets and foraging behaviour in baboons. Evolutionary Ecology 16: 579–593.

Howell DC (2002) Statistical methods for psychology (5th ed.). Duxbury Thompson Learning, Pacific Grove, CA, 384–387 pp.

Jenness J (2005) Distance Matrix (dist_mat_jen.avx) extension for ArcView 3.x, v. 2. Jenness Enterprises. Available at: http://www.jennessent.com/arcview/dist_matrix.htm

Kingdon J (1988) What are face patterns and do they contribute to reproductive isolation in guenons? In: A primate radiation: evolutionary biology of the African guenons (Gautier-Hion A, Bourliere F, Gautier J-P, Kingdon J, eds.), Cambridge University Press, Cambridge, England.

Kingdon J (1990) Island Africa: The evolution of Africa's rare animals and plants. William Collins Sons and Co. Ltd, London.

Kingdon J, Butynski TM, De Jong, Y (2008) Papio anubis. In: IUCN 2008. 2008 IUCN Red List of Threatened Species. <www.iucnredlist.org>. Downloaded on 27 March 2009.

Klingenberg CP, Barluenga M, Meyer A (2002) Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution 56: 1909–1920.

Kissling WD, Carl G (2008) Spatial autocorrelation and the selection of simultaneous autoregressive models. Global Ecology and Biogeography 17: 59–71.

Lawes MJ (1990) The distribution of the Samango monkey (Cercopithecus mitis erythrarchus Peters, 1852 and Cercopithecus mitis labiatus I. Geoffroy, 1843) and forest history in southern Africa. Journal of Biogeography 17: 669–680.

Lawes MJ (2002) Conservation of fragmented populations of Cercopithecus mitis in South Africa: the role of reintroduction, corridors and metapopulation ecology. In: The guenons: diversity and adaptation in African monkeys (Glenn ME, Cords M, eds.), Kluwer Academic/Plenum Publishers, New York, 375–392 pp.

Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? Ecology 74: 1659–1673.

Legendre P, Legendre L (1998) Numerical ecology (3rd ed.) Elsevier, Amsterdan, 853 pp.

Lomolino MV (2005) Body size evolution in insular vertebrates: generality of the island rule. Journal of Biogeography 32: 1683–1699.

Manly BFJ (1997) Randomization, bootstrap, and Monte Carlo methods in biology. Chapman and Hall, London.

Marroig G, Cheverud JM (2004) Cranial evolution in sakis (Pithecia, Platyrrhini) I: interspecific differentiation and allometric patterns. American Journal of Physical Anthropology 125(3): 266–278

Martins EP (2004) COMPARE, version 4.6b. Computer programs for the statistical analysis of comparative data. Distributed by the author at http://compare.bio.indiana.edu/. Department of Biology, Indiana University, Bloomington, IN.

Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. American Naturalist 149: 646–667.

Mayr R (1963) Animal species and evolution. The Belknap Press of Harvard University Press, Cambridge, MA.

Mazerolle MJ (2004) Making sense out of Akaike's Information Criterion (AIC): its use and interpretation in model selection and inference from ecological data. Thesis, Appendix 1, http://www.theses.ulaval.ca/2004/21842/apa.html

Napier PH (1981) Catalogue of primates in the British museum (Natural History) and elsewhere in the British isles. Part II: Family Cercopithecidae, Subfamily Cercopithecinae. British Museum (Natural History), London.

Nowak K, Cardini A, Elton S (2008) Evolutionary acceleration in an endangered African primate: speciation and divergence in the Zanzibar Red Colobus (Primates, Colobinae). International Journal of Primatology DOI 10.1007/s10764-008-9306-1.

Oates JF, Groves CP (2008) Cercopithecus nictitans. In: IUCN 2008. 2008 IUCN Red List of Threatened Species. <www.iucnredlist.org>. Downloaded on 25 March 2009.

Oates JF, Struhsaker T, McGraw S, Galat-Luong A, Galat G, Ting T (2008) Procolobus badius. In: IUCN 2008. 2008 IUCN Red List of Threatened Species. <www.iucnredlist.org>. Downloaded on 27 March 2009.

O'Higgins P, Jones N (2006) Tools for statistical shape analysis. Hull York Medical School, Hull/York.

Parenti LR, Humphries CJ (2004) Historical biogeography, the natural science. Taxon 53(4): 899–903

Peterson AT, Soberón J, Sánchez-Cordero V (1999) Conservatism of ecological niches in evolutionary time. Science 285: 1265–1267.

Polly PD (2008) Adaptive zones and the pinniped ankle: a 3D quantitative analysis of carnivoran tarsal evolution. In: Mammalian evolutionary morphology: a tribute to Frederick S. Szalay (Sargis E, Dagosto M eds.), Springer, Dordrecht, The Netherlands, 165–194 pp.

Rangel TFLVB, Diniz-Filho JAF, Bini LM (2006) Towards an integrated computational tool for spatial analysis in macroecology and biogeography. Global Ecology and Biogeography 15: 321–327.

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR (2005) Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. Journal of Human Evolution 48: 237–257.

Rohlf FJ, Marcus LF (1993) A revolution in morphometrics. Trends in Ecology and Evolution 8: 129–132.

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic Zoology 39: 40–59.

Rohlf FJ (2009) NTSYSpc. 2.20 V. ed. Setauket, Exeter Software, New York.

Ruggiero A, Kitzberger T (2004) Environmental correlates of mammal species richness in South America: effects of spatial structure, taxonomy and geographic range. Ecography 27: 401–416.

Rychlik L, Ramalhinho G, Polly PD (2006) Response to competition and environmental factors: skull, mandible, and tooth shape in Polish Water shrews (Neomys, Soricidae, Mammalia). Journal of Zoological Systematics and Evolutionary Research 44: 339–351.

Sanfilippo P, Cardini A, Mackey D, Hewitt A, Crowston J (2009) Optic disc morphology – rethinking shape. Progress in Retinal and Eye Research, 28: 227–248.

Slice DE (1999) Morpheus. beta ed. New York.

SPSS. 2006. SPSS for windows. 15.0 ed. SPSS Inc., Chicago.

Ting N (2008) Mitochondrial relationships and divergence dates of the African colobines: evidence of Miocene origins for the living colobus monkeys. Journal of Human Evolutione 55: 312–325.

Tosi AJ, Melnick DJ, Disotell TR (2005) Sex chromosome phylogenetics indicate a single tran-
   sition to terrestriality in the guenons (tribe Cercopithecini). Journal of Human Evolution 46:
   223–237.
Vitzthum VJ (1990) Odontometric variation within and between taxonomic levels of cercopitheci-
   dae: Implications for interpetations of fossil samples. Human Evolution 5: 359–374.
Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) Geometric morphometrics for biologists:
   a primer. Elsevier Academic Press, San Diego.

Chapter 9
Stock Identification of Marine Populations

Steven X. Cadrin

Idea and Aims

This chapter is intended to introduce researchers to an application of morphometric analysis that examines population structure of marine species. The approach has been effective for identifying the appropriate spatial scale for resource management units that reflect the underlying spatial patterns in the population. Several examples of stock identification of fishery resources are provided.

Introduction

Morphometric analysis is a valuable tool for scientists involved in studying marine populations, because it helps to identify intraspecific groups of animals that can be effectively monitored and conserved. Among a wide variety of methodological approaches, geographic patterns in morphology provide a unique perspective on spatial population structure. Conventions for sampling, analysis and interpretation have been developed to promote representative and meaningful conclusions, and many case studies have demonstrated the value of morphometric analysis for stock identification and resource conservation.

Management of living marine resources depends on accurate stock identification. Recognizing distinct subpopulations within a species is necessary to evaluate population trends and productivity as well as sustainability of human impacts. Population dynamics are more easily modeled and monitored if the appropriate "unit stock" is defined so that each of the components of biological production (growth, mortality and reproduction; Russell 1931) is determined from within the group rather than from outside the group. For example, new individuals enter a unit stock primarily from reproduction within the stock rather than from migrants from an adjacent

S.X. Cadrin (✉)

NOAA/UMass Cooperative Marine Education & Research Program, School for Marine Science, Science and Technology, 200 Mill Road, Suite 325, Fairhaven, MA 02719, USA
e-mail: Steven.Cadrin@noaa.gov

stock. Self-sustaining stocks are the units that resource managers need to recognize to meet their conservation and sustainability objectives.

Stock identification is a critical requirement of both conservation biology and fisheries science. Although the two fields have different objectives, they both involve inferences about how populations respond to perturbations or restoration efforts. Fishery management is usually focused on maintaining maximum sustainable yield (MSY; Mace 2001). Determination of MSY is most appropriate for demographically independent units that are essentially isolated from other groups on ecological time scales (i.e., years to decades). Conservation biology considers both long-term and short-term dynamics to determine the appropriate stock units to manage. Evaluating the risk of extinction for a species involves the recognition of "evolutionarily significant units" (Waples 1991) that are reproductively isolated over geologic time scales (i.e., millennia) and have developed unique adaptations. However, recovery of threatened species involves restoration of smaller, demographically independent management units that are self-sustaining on ecological time scales. Therefore, identification of unit stocks that are isolated from other stocks and maintain homogeneous vital rates is essential for management of living marine resources, whether for conservation biology or fishery management (Eagle et al. 2008).

Conservation biologists and fishery scientists approach the identification of stocks from many methodological perspectives to attain a holistic view of reproductive isolation and homogeneity (Begg and Waldman 1999). Stock identification has developed into an interdisciplinary synthesis that involves geographic distribution, movement among areas and geographic variation in genetics or phenotypic traits (Cadrin et al. 2005). Morphological variation is phenotypic (i.e., it is influenced by both genetic composition and environmental factors), and heritability of morphometric characters is generally low to moderate (Swain et al. 2005). Therefore, temporal stability in geographic variation of morphology is essential for stock identification applications, so that stocks are not determined on the basis of ephemeral differences in the environment. Despite the influence of environmental factors on morphological variation, patterns of variation can indicate groups that are isolated enough to maintain phenotypic differences. Morphometric patterns are also often associated with geographic differences in growth, maturity, or mortality which are critical to population dynamics. Adaptive significance of morphological features can add powerful interpretability to patterns of variation. For example, body form, fin size and fin location are adaptive for movement and maneuverability of fishes (Webb 1984) and cetaceans (Fish 1998); Position of the mouth and head shape are associated with trophic ecology (e.g., Costa and Cataudella 2007); Abdomen size reflects energetic investment, feeding and spawning condition (e.g., Armstrong and Cadrin 2001); Body armor (e.g., scutes, spines) can indicate different predatory environments (e.g., Walker 1996); and secondary sex characters (e.g., dorsal humps and fin size of fishes; chelipeds of crustaceans) can indicate behavioral differences and size at maturity (Cadrin 2000).

Morphometric analysis offers a unique perspective to the investigation of population structure. Patterns in morphology have been used to identify and discriminate stocks for nearly a century (Teissier 1936). More recently, morphometric patterns

are considered in the context of information on genetic variation and movement patterns for an interdisciplinary analysis of population structure.

Methods

Conventional protocols for sampling and analyzing morphometric data have been developed for a wide range of applications. Methods for morphometric stock identification reflect general conventions, and additional features should be considered that are more specific to investigating stock structure.

Sampling Designs

Stock identification typically progresses in stages, from exploratory to confirmatory, and is finally applied as either classification or mixture analysis. Each stage has different sampling requirements:

Exploratory Stock Identification – When little is known about stock structure of a species, general information on geographic distribution (e.g., research surveys, fishing grounds) should be used to define putative stocks, and each of them should be sampled to explore general patterns of variability throughout the range of the species. At this stage, wide geographic coverage of samples across areas and seasons is more important than sample sizes within groups, because groups have not been defined. Optimal sample sizes should be based on stability of pooled-group multivariate analyses (e.g., there should be at least three times as many samples as variables; Saila and Martin 1987). At the exploratory stage, data collection should include as many potentially explanatory variables as possible (e.g., age, gender, maturity stage, depth, color) to be considered as potential factors to be included in analysis of morphometric patterns. As with any study, a comprehensive literature review is valuable at the exploratory stage to design sampling.

Confirmatory Stock Identification – Once a stock structure hypothesis has been developed, either by exploratory analysis of morphometric variation or information from other approaches to stock identification, "baseline samples" from each putative stock should be collected in areas and at a time of year when stock mixing is minimal (e.g., during the spawning season, on spawning grounds). Sample sizes within groups should be a minimum of 50 specimens to derive reliable estimates of covariance among features (Tabachnick and Fidell 1989), recognizing that sexual dimorphism may require separate-sex analyses. A critical aspect of sampling design for confirmatory analysis is to include all putative stocks of interest. Including an "out-group" (i.e., another group far from the area of interest) can provide context in which to understand the magnitude of variation among the more related groups being studied.

Stock Discrimination – Once substantial and meaningful differences are confirmed to exist among stocks, those differences can be used to classify individuals to a stock. Accordingly, areas between the source samples used for confirmatory

analysis (i.e., areas and seasons with possible mixed stocks) should be representatively sampled. A statistically significant difference among baseline samples is essential, but not always enough, for accurate classification. Although there is no rule of thumb to determine the magnitude of difference needed for accurate discrimination, there are useful diagnostics to judge performance (e.g., extrinsic classification or cross-validation). Stock discrimination can be used to help delineate geographic (and possibly seasonal) boundaries among stocks or to determine stock composition in a mixture (e.g., a mixed-stock fishery).

Processing Specimens

Many aspects of image processing for stock identification are similar to other applications of morphometric analysis. For example, images must be calibrated, measurement error should be calculated over repeated measures, and shape change associated with preservation method should be assessed. However, the choice of morphological features for possible stock identification characters may require additional consideration.

Most of the recent advances in morphometrics have been focused on landmarks (e.g., Elewa 2004), and landmark methods comprise a large portion of the research in morphometric stock identification. However, many studies use other aspects of morphology to study stock structure such as outline features, circuli patterns and meristics. All four categories of features can be measured using image analysis.

Landmark features can be used on any morphological feature (e.g., bones, otoliths), but are generally used to measure general body form. Some landmark features that are associated with biologically meaningful variation are fin placement, body depth, caudal depth, head size, position of the mouth, and eye position. Most of these features are captured by Straus and Bookstein's (1982) box-truss network, which was initially developed for fishes, but is applicable for diverse taxa. Although the box-truss was initially developed for traditional multivariate analysis of box-truss distances, the box-truss landmarks are also valuable for geometric analysis, because they represent an efficient measure of general body form that is related to function. An augmented set of triangle-truss landmarks and linear distances that capture general body form, fin size, mouth size and eye position is illustrated in Fig. 9.1.

Outline features have also been extensively applied to stock identification. Outlines of hard parts such as otoliths (i.e., "ear bones") or scales are typically used to investigate geographic variation of bony fishes, but perhaps the most biologically meaningful patterns of variation in outline shape have been found for bivalves. For example, ontogenetic stages of Atlantic sea scallop (*Placopecten magellanicus*) are reflected in outline shape of their valves (Dadswell and Weihs 1990). After juvenile scallops detach themselves from their byssal threads, they are active swimmers until they become more sedentary adults. During the juvenile, swimming phase, the valve hinge becomes more prominent, and the valve (shell) becomes more circular, both of which confer greater swimming velocity and duration; during the adult, sedentary phase, the valve becomes wider and more oval (Fig. 9.2, C. O'Keefe SMAST video
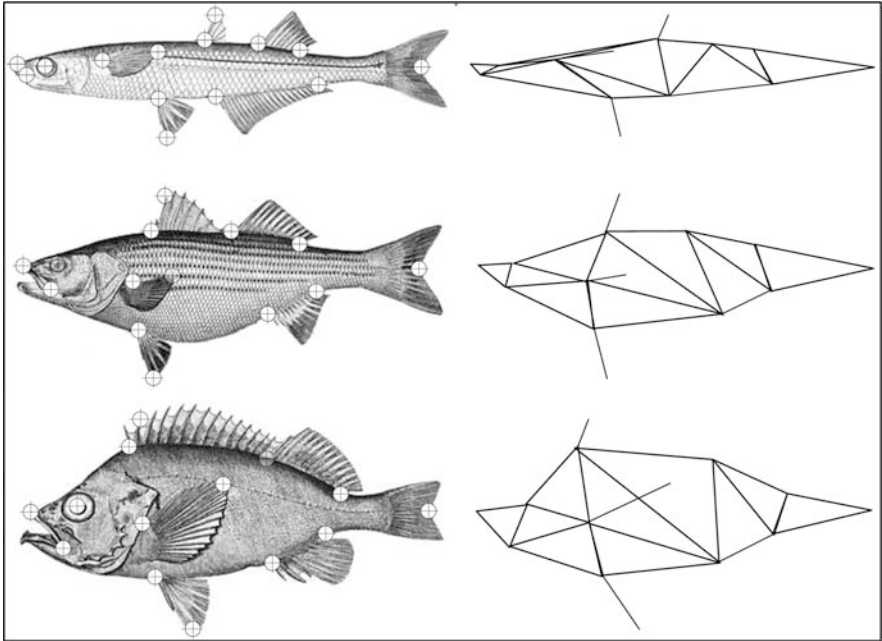
Fig. 9.1 Augmented triangle truss networks of landmarks ("targets" on fish images) and distances (line segments on *right*) for Atlantic silverside (*Menidia menidia*), striped bass (*Morone saxatilis*) and Acadian redfish (*Sebastes fasciatus*); fish drawings from Bigelow and Schroeder (1953)
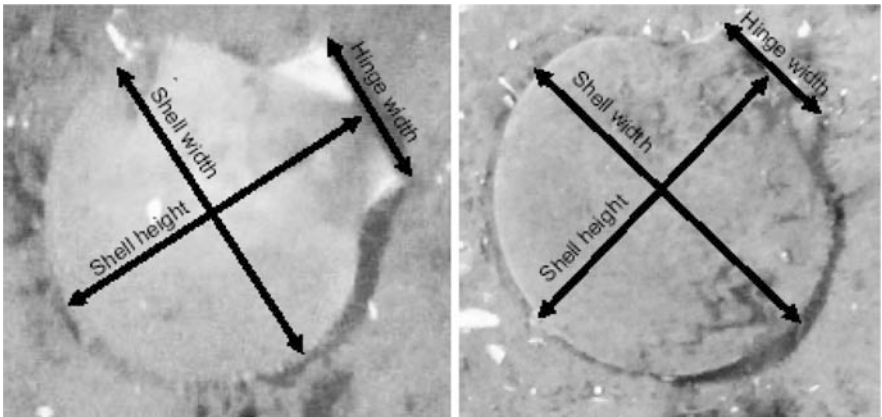
Fig. 9.2 Morphometry of sea scallop (*Placopecten magellanicus*) shells. Note that the juvenile *(left)* has nearly equal shell height and shell width and a wider relative hinge width than the adult *(right)* which is has greater shell width than shell height (images form C. O'Keefe, SMAST video survey)

survey). Kenchington and Full (1994) quantified these ontogenetic shape changes using Fourier analysis of valve outlines to discriminate four scallop beds. The geographic variation in valve shape was associated with differences in age at maturity and swimming behavior.

Landmark and outline features are commonly used characters for morphometric analysis and morphometric stock identification. Two other aspects of morphology, circuli patterns and meristics, have also been useful for stock identification. Circuli patterns are growth rings in the hard parts. Otoliths of bony fishes are used for a wide range of applications to identify annuli (i.e., yearly bands) to determine age, to identify daily growth rings to determine hatch date, and to investigate differences in growth rates among groups. For example, relative growth rates of Atlantic salmon (*Salmo salar*) can be measured from circuli patterns on scales (Fig. 9.3, F. Hogan,
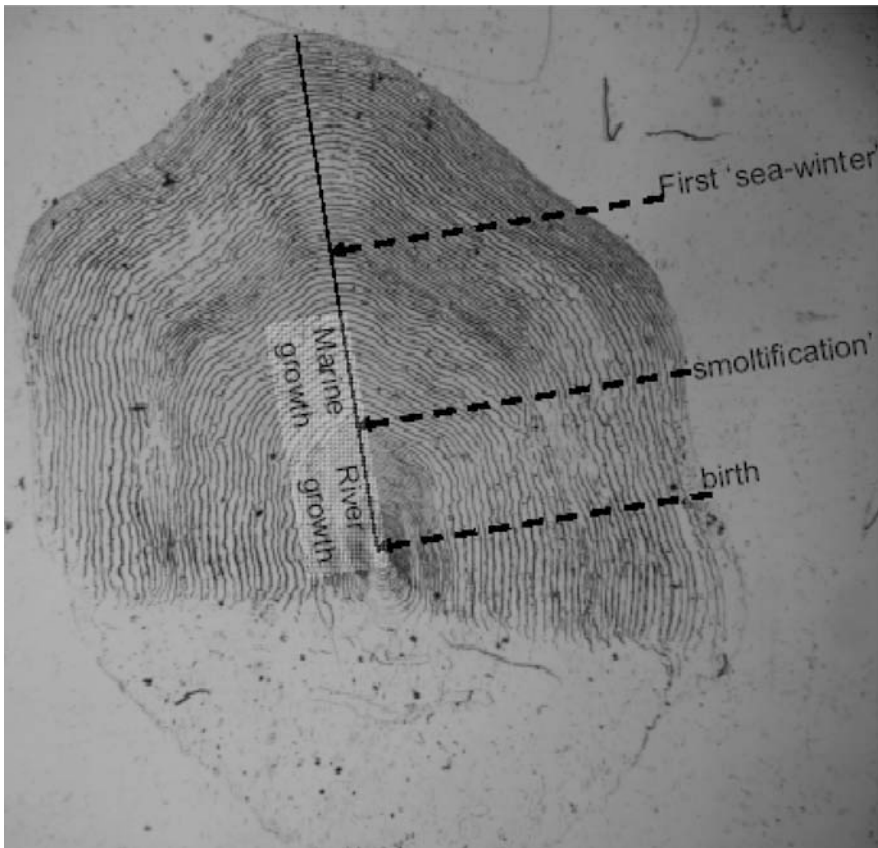
Fig. 9.3 Image of a scale from Atlantic salmon (*Salmo salar*). The solid line is a 'transect' from the birthmark to the scale margin, *arrows* indicate ontogenetic stages: birth, 'smoltification' (i.e., the transition form river to marine environments), and the first winter in the marine environment (image form F. Hogan, NOAA/UMass CMER Program)

University of Massachusetts) and compared among stocks. The advantage of using circuli patterns to study stock structure is that they are directly related to growth rate, a phenotypic character that is critical in modeling population dynamics and sustainability.

Meristics are discrete morphological features that can be counted, rather than measured (e.g., vertebrae, fin rays, gill rakers). Meristic characters have been used extensively for stock identification as a phenotypic character that is typically influenced by early-life history environments. Merisitic values (i.e., the number of discrete parts) are often determined during a distinct ontogenetic stage when the features are developing and sensitive to environmental influences. For example, the number of fish vertebrae is determined during the larval period, and is usually inversely related to temperature (Lindsey 1988).

Similar to the way morphometric analysis of landmarks and outlines have benefited from technological developments in image analysis, the study of circuli patterns and meristics have advanced in recent decades. Circuli patterns, including the number of circuli and circuli spacing, can be measured using luminescence transects from the focus or center of the growth structure to the margin (Fig. 9.3). Meristics can also be counted via imaging.

All four categories of features (landmarks, outlines, circuli or meristics) rely on thoughtful choice of characters to efficiently represent morphology with the fewest number of characters. There are an infinite number of morphometric features that can be measured, many of which are redundant, some of which are critical for capturing meaningful variation among stocks. Homology of characters (i.e., the same developmental feature among specimens; Strauss and Bookstein 1982) is critical for standardization and supports biological interpretations. Measurement of characters must also be repeatable among specimens, researchers and methodologies. Finally, prior information on life history and morphological patterns in related species should be considered in the choice of characters to measure.

Statistical Analysis

Morphometric data are multivariate in the strictest sense: several to many variables are needed to measure morphometry, and each variable is structurally correlated with others. All dimensions increase with size, and shape variation is a contrast in relative growth. Therefore, all categories of morphometric data are analyzed using multivariate analysis.

A simple example of a hypothetical sample of nine lumpfish (*Cyclopterus lumpus*) specimens is illustrated in Fig. 9.4 to demonstrate how morphometric data can be analyzed using multivariate analysis. The nine lumpfish represent a range of sizes and shapes that are crudely measured using a simple box truss network of six distances.

The data measured from the hypothetical lumpfish are listed in Table 9.1 as raw distances (cm) and natural log transformed distances. Log transformation is typical in morphometric analysis for statistical reasons (bivariate relationships are more
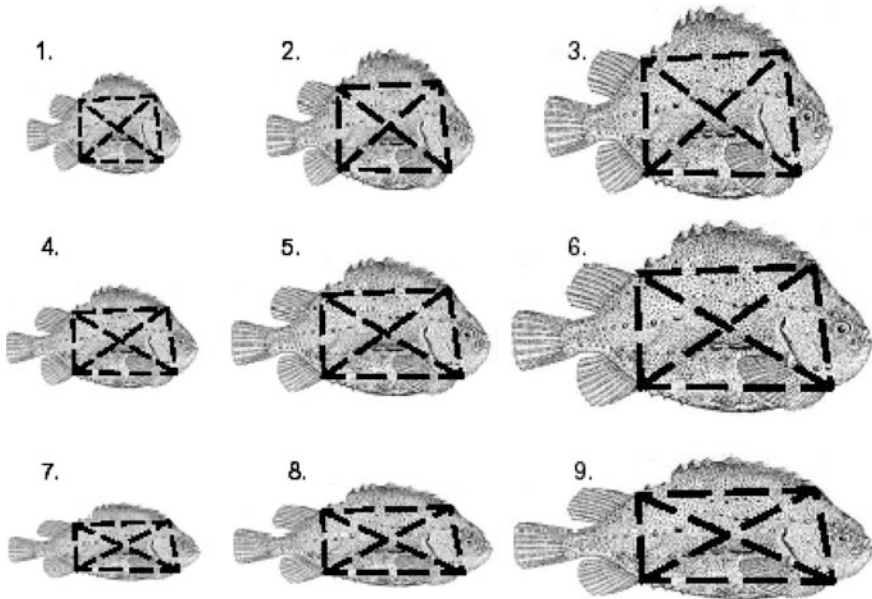
Fig. 9.4 Hypothetical sample of nine lumpfish, with a simple box-truss network of linear morphometric distances

linear, variances are more homogeneous) as well as theoretical reasons (distances increase by multiplicative growth; Huxley 1932). Note that log transformed data have more similar arithmetic means and standard deviations than the untransformed data. The six variables measured from the lumpfish specimens are highly correlated (Fig. 9.5).

Principle components analysis (PCA) shows that the greatest source of variation and covariation in the data is related to size (Table 9.2), because all six variables "load positively" on the first principle component (i.e., each variable is positively related to a composite variable that explains the most variation in the data). The second component of variance is a shape component, because it contrasts body depth (i.e., head depth and caudal depth load strongly positive) and body length (i.e., dorsal length and ventral length load strongly negative; diagonal distances load moderately negative). Therefore, most variance in the data (PC1) is from size differences among specimens, and the greatest variance in shape (PC2) is differences in relative depth.

The distribution of PC scores for each specimen also shows that large fish have positive PC1 scores, and small fish have negative PC1 scores (Fig. 9.6). The shape axis (PC2) clearly distinguishes deep and short fish (those with positive PC2 scores) from shallow and long fish (those with negative PC2 scores). There are three groups of fish with different shape: deep-short, intermediate and shallow-long; each group has three specimens: small, intermediate and large.

This simple example shows how multivariate analysis can examine patterns of variation for several correlated variables by deriving two composite variables that

Table 9.1 Six linear distances measured from nine hypothetical lumpfish (untransformed data in cm, above; log transformed data below)

Specimen	Head depth	Caudal depth	Dorsal length	Ventral length	Diagonal 1	Diagonal 2
1	9	8	12	13	20	18
2	16	14	20	22	31	29
3	26	23	31	34	46	43
4	9	8	17	20	25	23
5	16	14	27	30	37	35
6	26	23	42	45	56	52
7	5	4	17	20	23	20
8	11	8	27	30	35	31
9	17	15	42	45	52	46
Mean	15.0	13.0	26.1	28.8	36.1	33.0
Std.dev.	7.4	6.7	10.8	11.2	12.9	12.0
1	2.20	2.08	2.48	2.56	3.00	2.89
2	2.77	2.64	3.00	3.09	3.43	3.37
3	3.26	3.14	3.43	3.53	3.83	3.76
4	2.20	2.08	2.83	3.00	3.22	3.14
5	2.77	2.64	3.30	3.40	3.61	3.56
6	3.26	3.14	3.74	3.81	4.03	3.95
7	1.61	1.39	2.83	3.00	3.14	3.00
8	2.40	2.08	3.30	3.40	3.56	3.43
9	2.83	2.71	3.74	3.81	3.95	3.83
Mean	2.59	2.43	3.18	3.29	3.53	3.44
Std.dev.	0.54	0.57	0.43	0.41	0.37	0.38
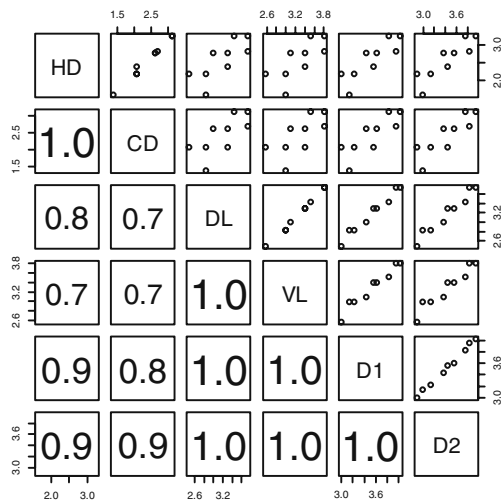
Fig. 9.5 Bivariate relationships among the six box-truss variables measured from the lumpfish sample. Data are natural log transformed. Variable codes are on the diagonal, bivariate distributions are above the diagonal and correlation coefficients are *below*

Table 9.2 Principle component (PC) loadings for six linear distances measured from nine hypothetical lumpfish (data in Table 9.1)

Component	Head depth	Caudal depth	Dorsal length	Ventral length	Diagonal 1	Diagonal 2
PC1	0.490	0.509	0.374	0.351	0.337	0.352
PC2	0.427	0.537	− 0.458	− 0.484	− 0.232	− 0.180

can be more simply interpreted. For morphometric stock identification, the first principle component of variance is typically removed to derive size-free shape discrimination among groups (e.g., Burnaby 1966). If the deep-short lumpfish were from one putative stock, and shallow-long fish were from another, a size-adjusted discriminant function would be similar to PC2.

Once a discriminant function has been developed from a sample of known group membership (i.e., "baseline samples"), the function can be used to classify individual specimens to a stock (i.e., "mixture samples"). One measure of classification accuracy is the ability to classify baseline samples, but this intrinsic classification tends to overestimate accuracy, because it is based on the same specimens that are used to establish discriminant functions. A more reliable measure of accuracy is
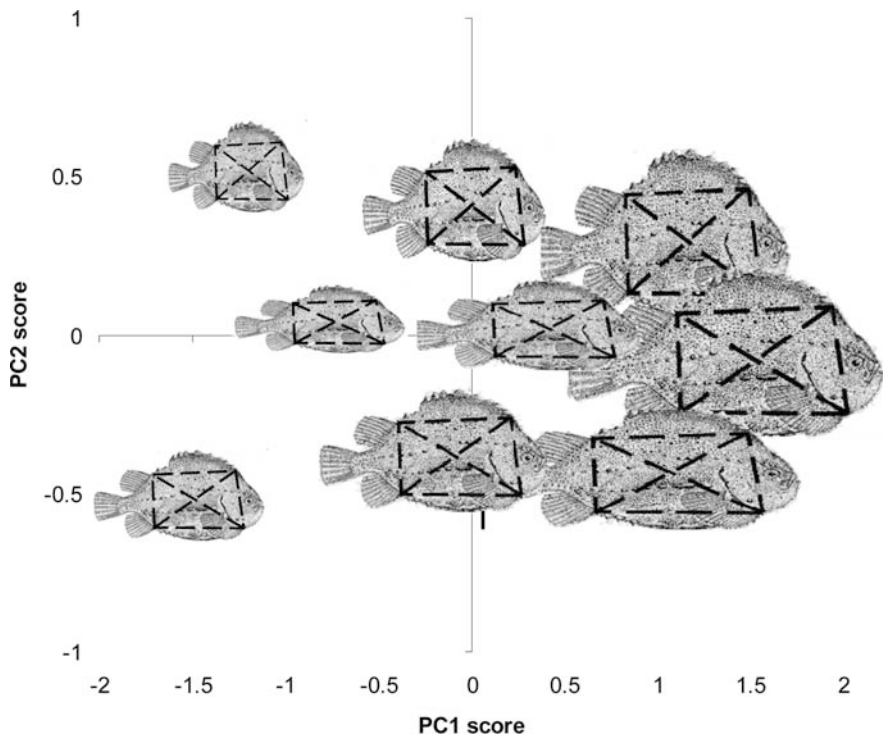
Fig. 9.6 Principle component scores of the nine lumpfish specimens (images age centered on the specimen's score), with PC1 serving as an index of size, and PC2 an index of relative body depth

extrinsic classification, or cross-validation, which uses known-stock individuals that are not in the baseline sample to verify accurate classification. Discriminant scores can be used to delineate stock boundaries or to determine the composition of each stock in a mixture. For example, stock composition can be determined for the catch of a fishery that exploits a mixture of spawning groups during the feeding season.

The lumpfish example above used traditional linear distances, but data from outlines (e.g., amplitudes from a series of Fourier harmonics) or circuli analyses (e.g., distances between successive circuli) can be similarly analyzed with PCA and discriminant analysis. One disadvantage of multivariate analysis of traditional linear distances is that it ignores the relative geometry of dimensions (e.g., head depth and dorsal length of lumpfish share a landmark). Geometric analysis of landmarks retains the relative position of landmarks and quantifies deformations from one set of landmarks to another. For example, Fig. 9.7 depicts the flattening needed to
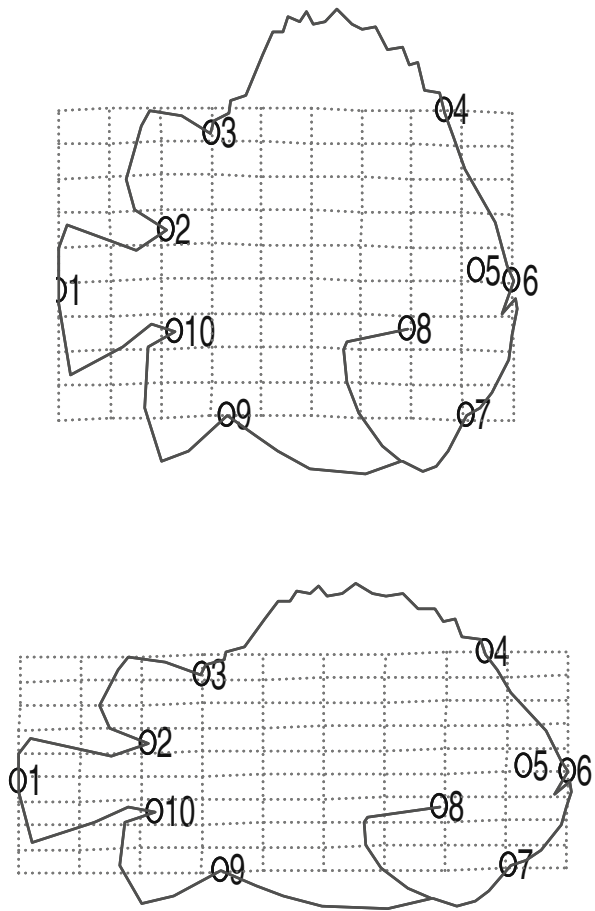
Fig. 9.7 Deformation from lumpfish specimen 1 to specimen 7, as measured by thin plate spline analysis (landmarks are numbered, the outline and grid are overlaid to help interpret shape variation)
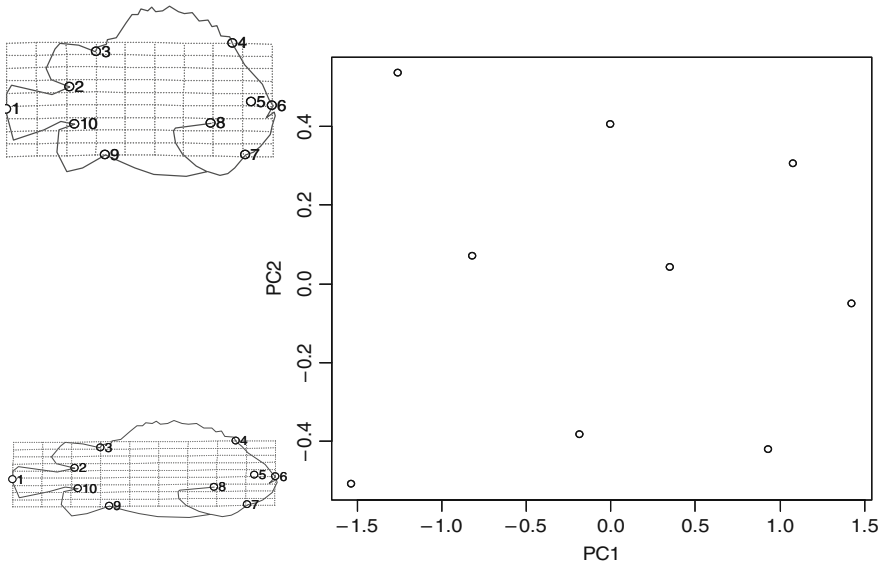
Fig. 9.8 Synthetic approach of traditional multivariate analysis of linear box-truss distances (PC score plot on *right*) and geometric analysis of landmarks (fish images on *left*), as illustrated using PC2 of the lumpfish data, which is associated with a general flattening

transform lumpfish specimen 1 to specimen 7. Similar to traditional distances, geometric deformations can also be used for morphometric stock identification.

Several recent case studies of morphometric stock identification have combined traditional analyses of linear distances with geometric analysis of landmarks (Cadrin and Silva 2005; Sheehan et al. 2005). The synthetic approach involves multivariate analysis of linear dimensions and geometric analysis of extreme specimens to illustrate the deformations associated with components of variation or differences among groups. For example, the geometric analysis of the extreme PC2 scores from the lumpfish data shows that PC2 is associated with a flattening (Fig. 9.8).

Conclusions

Morphometric analysis provides a complementary tool for stock identification. It can reveal patterns of life history that vary among subpopulations, indicating limited movement between groups and possible genetic variation. Therefore, in concert with information on genetic isolation, movement patterns and other phenotypic variation, morphometry has a distinct role in stock identification of marine populations. One of the strengths of morphometric stock identification is that patterns of geographic variation can be interpreted in terms of adaptive characters that influence population dynamics. Thoughtful interpretation of morphometric patterns confers meaning and emphasizes unique role of morphometrics in interdisciplinary analyses.

The value of morphometric analysis for determining stock structure relies on appropriate methodology. Well designed sampling and analysis supports representativeness and reliability of results. As a distinct application of morphometric analysis, case studies in morphological stock identification offer methodological advancements and refinements that could be considered in other morphometric applications.

References

Armstrong MP, Cadrin SX (2001) Morphometric patterns within and among spawning aggregations of Atlantic herring (*Clupea harengus*) off the northeast United States. pp 575–590 In Herring 2000, F. Funk, J. Blackburn, D. Hay, A.J. Paul, R. Stephenson, R. Toressen and D. Witherell, eds. Alaska Sea Grant Report AK-SG-2000-01.

Begg GA, Waldman JR (1999) A holistic approach to fish stock identification. Fish Res 43: 37–46.

Bigelow HB, Schroeder WC (1953) Fishes of the Gulf of Maine. Fishery Bulletin of the Fish and Wildlife Service Vol. 53.

Burnaby TP (1966) Growth-invariant discriminant functions and generalized distances. Biometrics 22: 96–110

Cadrin SX (2000) Advances in morphometric analysis of fish stock structure. Rev Fish Biol Fish 10: 91–112.

Cadrin SX, Friedland KD, Waldman J (2005) Stock identification methods: applications in fishery science. Elsevier Academic Press, San Diego.

Cadrin SX, Silva VM (2005) Morphometric variation of yellowtail flounder. ICES J Mar Sci 62: 683–694.

Costa C, Cataudella S (2007) Relationship between shape and trophic ecology of selected species of Sparids of the Caprolace coastal lagoon (Central Tyrrhenian sea). Environ Biol Fish 78: 115–123.

Dadswell MJ, Weihs D (1990) Size-related hydrodynamic characteristics of the giant scallop, *Placopecten magellanicus* (Bivalvia: Pectinidae). Can J Zool 68: 778–785.

Eagle TC, Cadrin SX, Caldwell ME, Methot RD, Nammack MF (2008) Conservation units of managed fish, threatened or endangered species, and marine mammals. NOAA Technical Memorandum NMFS-OPR-37.

Elewa AMT, ed. (2004) Morphometrics–applications in biology and paleontology. Springer-Verlag Publishers, Heidelberg, Germany. 263 pp.

Fish FE (1998) Comparative kinematics and hydrodynamics of odontocete cetaceans: morphological and ecological correlates with swimming performance. J Exp Biol 201: 2867–2877.

Huxley JS (1932) Problems of relative growth. The Dial Press, New York. 276 pp.

Kenchington EL, Full WE (1994) Fourier analysis of sea scallop (*Placopecten magellanicus*) Shells in determining population structure. Can J Fish Aqua Sci 51: 348–356.

Lindsey CC (1988) Factors controlling meristic variation. pp 197–274 In Fish Physiology, Viviparity and Posthatching juveniles, W.S. Hoar, and D.J. Randall, eds. Academic Press.

Mace PM (2001) A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management. Fish Fish 2: 2–32.

Russell ES (1931) Some theoretical considerations on the "overfishing" problem. Cons Perm Int Explor Mer 6: 3–20.

Saila SB, Martin BK (1987) A brief review and guide to some multivariate methods for stock identification. pp 149–175 In Proceedings of the Stock Identification Workshop, H.E. Kumpf, R.N. Vaught, C.B. Grimes, A.G. Johnson, and E.L. Nakamura, eds. NOAA Tech. Mem. NMFS-SEFC 199.

Sheehan TF, Kocik JF, Cadrin SX, Legault CM, Atkinson E, Bengtson D (2005) Marine growth and morphometrics for three populations of Atlantic salmon from Eastern Maine, USA. Trans Am Fish Soc 134: 775–788.

Strauss RE, Bookstein FL (1982) The truss: body form reconstructions in morphometrics. Syst Zool 31: 113–135.

Swain DP, Hutchings JA, Foote CJ (2005) Environmental and genetic influences on stock identification characters. pp 45–85 In Stock Identification Methods, S.X. Cadrin, K.D. Friedland and J.Waldman, eds. Elsevier Academic Press, San Diego.

Tabachnick BG, Fidell LS (1989) Using multivariate statistics. Harper Row and Collins, New York. 746 p.

Teissier G (1936). Croissance comparee des formes locales d'une meme espece. Mem Mus r Hist Nat Belg 3 (2nd Ser): 627–634.

Webb PW (1984) Form and function in fish swimming. Sci Am 251: 72–82.

Walker JA (1996) Principal components of body shape variation within an endemic radiation of threespine stickleback. pp 321–334. In Advances in Morphometrics. L.F. Marcus, M. Corti, A. Loy, G.J.P. Naylor, and D.E. Slice, eds. NATO ASI Series A: Life Sciences 284.

Waples RS (1991) "Definition of 'Species' Under the Endangered Species Act: Application to Pacific Salmon." U.S. Department of Commerce. NOAA Technical Memorandum NMFS F/NWC-194.

Chapter 10
Correlating Shape Variation with Feeding Performance to Test for Adaptive Divergence in Recently Invading Stickleback Populations from Swiss peri-alpine Environments

Denis Roy, Kay Lucek, Esther Bühler, and Ole Seehausen

Idea and Aims

The purpose of this chapter is to demonstrate the application of geometric morphometrics in a typical study, and put the information it provides into a broader context. Here we use geometric morphometrics to describe the head shape among three different Swiss stickleback populations from two drainages, including both lake and stream residents. Head shapes are compared to feeding efficiency indices generated from laboratory trials using lake and stream prey types. We also combine these data with genetic and other more traditional morphological assessments to understand the roots of the tremendous variation exhibited by sticklebacks in Switzerland. This work shows that in combination with other data, geometric morphometrics can make a significant contribution toward understanding the natural history of taxa and is an indispensible tool providing insight into fundamental mechanisms of adaptive divergence and speciation.

Introduction

A fundamental tenet of adaptive radiation based on feeding resources is that trophic morphology expressed by divergent populations, or closely related species, is related to the food resources available within the local environment (Schluter 1995, 2000; Coyne and Orr 2004). A correlation between phenotype and environment is a defining characteristic of adaptive radiations as expressed by both Schluter (2000) and Bernatchez (2004). Although a correlation between morphological variation and an environmental gradient is suggestive, to demonstrate that traits are indeed adaptive, they must also generate an advantage for their bearer in accessing resources, often

––––––––––––––––––

D. Roy (✉)
Biology Department, Dalhousie University, 1355 Oxford St. Halifax, NS B3H 4J1, Canada;
EAWAG, Swiss Federal Institute for Aquatic Sciences and Technology, Center of Ecology, Evolution & Biogeochemistry, Seestrasse 79, CH6047 Kastanienbaum, Switzerland
e-mail: denisroy1@gmail.com

referred to as "trait utility" (Schluter 2000; Bernatchez 2004; Coyne and Orr 2004; Rundle and Nosil 2005). In this context, populations adapted to different environments wherein differential selection has acted on feeding morphologies should not only express different morphologies, but these morphologies should also be more efficient in their respective feeding environments for exploiting available resources (Schluter 2000; Bernatchez 2004 Nosil and Reimchen 2005).

Many morphological traits have been linked to feeding and foraging efficiency, especially in fish (Skùlason et al. 1993; Schluter 1994; Bernatchez 2004; Roy et al. 2007). For example, Rincòn et al. (2007) demonstrated that size, position and orientation of the mouth, body depth, and position of the dorsal fin have a substantial impact on prey capture success in drift feeding cyprinids. Similarly, many studies infer differential feeding success of closely related species based on the density and lengths of gillrakers used to sieve and redirect plankton into the buccal cavity (e.g., Hudson et al. 2007; Vonlanthen et al. 2009). Thus, several morphological traits from gillrakers to pharyngeal jaw shape have been associated with feeding differences among closely related phenotypes and species (Albertson et al. 2003; Bernatchez 2004; Roy et al. 2007). Yet, despite many studies identifying key phenotype – environment correlations in taxa purported to exhibit adaptive radiation, few demonstrate the utility of these traits in accessing different resources (Schluter 2000; Rundle and Nosil 2005).

Threespine sticklebacks (*Gasterosteus aculeatus*) are an excellent model species for the study of evolutionary processes (Bell and Foster 1994; McKinnon and Rundle 2002). Many young ecologically divergent species exist and several instances have been reported of parallel adaptation in North American and European postglacial freshwater systems (Bell and Foster 1994; Schluter and Nagel 1995; McKinnon and Rundle 2002; Mäkinen et al. 2006). In many such systems, sticklebacks have diverged into lake-dwelling and stream-dwelling forms wherein the composition of invertebrate communities and environmental factors such as flowrate, water temperature and predation regimes are quite distinct (Reusch et al. 2001; Hendry et al. 2002; Hendry and Taylor 2004; Mäkinen et al. 2006; Moore et al. 2007; Berner et al. 2008). The selection pressures on traits related to food acquisition and use should therefore differ markedly between these two habitat types.

Lake and stream dwelling forms show divergent morphologies wherein lake forms typically have slimmer bodies and heads, long snouts with narrower more upturned mouths, larger more sunken eyes, and longer more numerous gillrakers (Hendry et al. 2002). These features appear well adapted for feeding on pelagic food sources such as copepods and Daphnids. Conversely, stream dwellers usually have stouter bodies and heads, shorter snouts equipped with larger more downturned mouths, smaller more elevated eyes and fewer shorter gillraker, features more typically associated with feeding on benthic macroinvertebrates (Walker 1997; Caldecutt and Adams 1998; Walker and Bell 2000; Hendry et al. 2002; Hendry and Taylor 2004; Berner et al. 2008). The different morphologies expressed by lake and stream sticklebacks are believed to be specifically adapted and thought to provide more efficient foraging in their respective environments.

Swiss sticklebacks, however, have been seldom studied mostly because of their purported absence from pre-alpine and alpine lakes and tributaries (Fatio 1882; Bertin 1927; Munzig 1963; Mäkinen et al. 2006; but see Milinski and Bakker 1990; Bakker 1993). Sticklebacks have a limited history in Swiss aquatic systems with initial reports dating to the 1870s of a restricted distribution located in the lower Rhine area near Basel (Fatio 1882). Near the turn of the century however, isolated populations were also identified in the upper Rhine region upstream of Lake Constance and in the Rhone drainage near Lakes Neuchatel and Geneva (Fatio 1882; Bertin 1927). Since their initial observations, sticklebacks have become widespread within Switzerland and have established many populations both within large peri-alpine lakes (e.g., Lake Constance, Lake Geneva) and in small lake tributaries and river systems. Currently, it is not known whether Swiss populations colonized through different rivers and lakes, if they spread from their initial Rhine population or if their distribution has been facilitated by multiple invasions.

Nevertheless, and despite their relatively recent history, Swiss sticklebacks exhibit tremendous phenotypic variation in many ecologically relevant traits including body armor, size and overall morphology (*personal observation*), similar to that observed over the much larger, continental European scale including the Sea (Munzig 1963), and to that observed over many North American postglacial lakes and rivers (Bell and Foster 1994; Taylor and McPhail 2000; Walker and Bell 2000; Hendry et al. 2002; Mäkinen et al. 2006). Sticklebacks are renowned for exhibiting rapid morphological evolution in adaptation to different environments (Walker and Bell 2000; Peichel et al. 2001; Kraak et al. 2001; Kitano et al. 2008; Barret et al. 2008). It is therefore conceivable that much of the morphological variation expressed in Swiss populations has accumulated quickly within the last 100 years (approximately 100+ generations) in response to divergent selection. Alternatively, differences between populations may have evolved outside their new Swiss distribution, representing historical contingency.

In this study, we use geometric morphometrics (GM) to quantify differences in head shape among three different Swiss stickleback populations, including one lake and two stream residents, taken from two drainage systems (Geneva and Constance). We focus on head shape differences because this is the primary area of the body associated with trophic traits influencing feeding efficiency (e.g., visual ability, mouth size and position, gillraker morphology). All fish characterized by GM were then tested in feeding efficiency trials using both lake and stream type food in a laboratory setting. This was done to see if population specific head shapes exhibited differential feeding efficiencies as expected based on the trait-utility hypothesis (Schluter 2000; Bernatchez 2004; Rundle and Nosil 2005). Phylogenetic relationships among sampled populations were determined using mtDNA sequence variation to clarify the historical context. The same three populations were also sampled for anti-predator defense traits to determine if divergent predation regimes may explain some of the substantial phenotypic variation.

We predict that lake and stream fish will exhibit head shapes commonly associated with lake and stream phenotypes as described above (Hendry et al. 2002; Berner et al. 2008). We also predict that if populations share recent common ancestry within

Switzerland, then head shape differences among them are likely to be adaptive, i.e., that divergent selection pressures between environments likely shaped their divergence. In this case, we predict that each shape will be more efficient at feeding on lake or stream type food, respectively. In contrast, no such prediction can be made if populations are distantly related because differences could reflect historical contingency as opposed to adaptive divergence in the context of their current environments. Alternatively, divergence among populations may be driven by other factors not necessarily linked to feeding efficiency. We show that GM can be a very useful tool clarifying the natural history of populations, especially when incorporated into a research framework that includes other corroborating evidence such as genetics and ecology (Rohlf and Marcus 1993; Adams et al. 2004; Rincòn et al. 2007; Roy et al. 2007).

Methods

Freshwater sticklebacks from three different populations were collected in late spring of 2007 from Lakes Constance and Geneva areas (Fig. 10.1). Lake Constance is part of the Rhine River system and the populations collected there included one
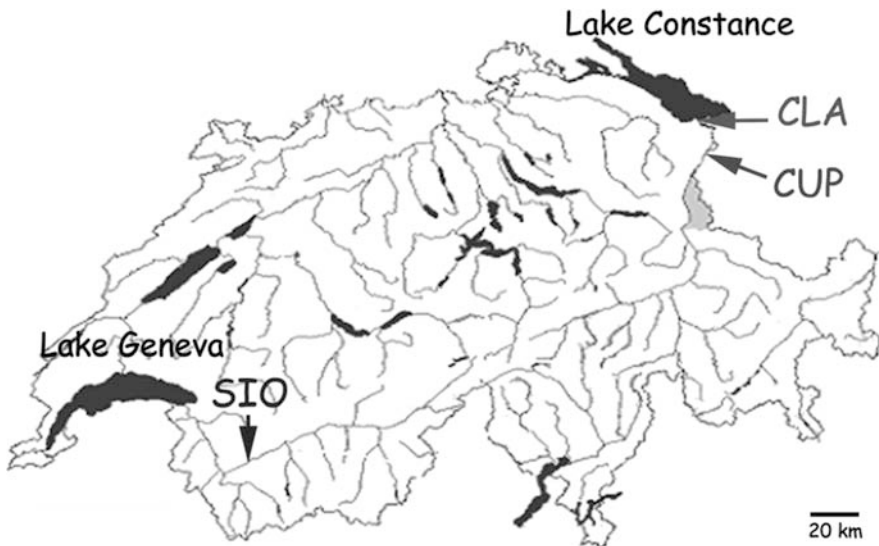
Fig. 10.1 Location of Sampling sites for Swiss stickleback populations. CLA = Lake Constance near Buriet (N47°29′02.3″, E9°33′35.3″), CUP = upper tributary of Lake Constance near Oberriet (N47°19′29.4″, E9°33′55.″) & SIO = Sion Canal in the Valais region (near Sion Airport: N46°12′50.6″, E7°18′53.0″)

from the lake proper (CLA) and another located ~24 km up one of the lake's tributaries (CUP). Lake Geneva, on the other hand, belongs to the Rhone River drainage and a resident stream population was collected from Sion Canal (SIO), an estimated ~60 km upstream from the lake (Fig. 10.1). Collected fish were placed in population-specific aerated containers and transported back to the laboratory where they were housed in flow through tanks maintained at 14–16°C.

Geometric Morphometrics

Upon capture, the head of 16 individual fish from each population was photographed in a standardized manner. Head shape differences among populations were quantified using geometric morphometrics (Rohlf 1990; Rohlf et al. 1996; Adams et al. 2004; Zelditch et al. 2004). Eight homologous landmarks were chosen corresponding to several important head features associated with both feeding and prey detection (Fig. 10.2) (see Hart and Gill 1994; Caldecutt and Adams 1998; Roy et al. 2007). The same person set all landmarks on all photos to avoid multi-user biases in landmark placement. Specimen landmark configurations were then subjected to
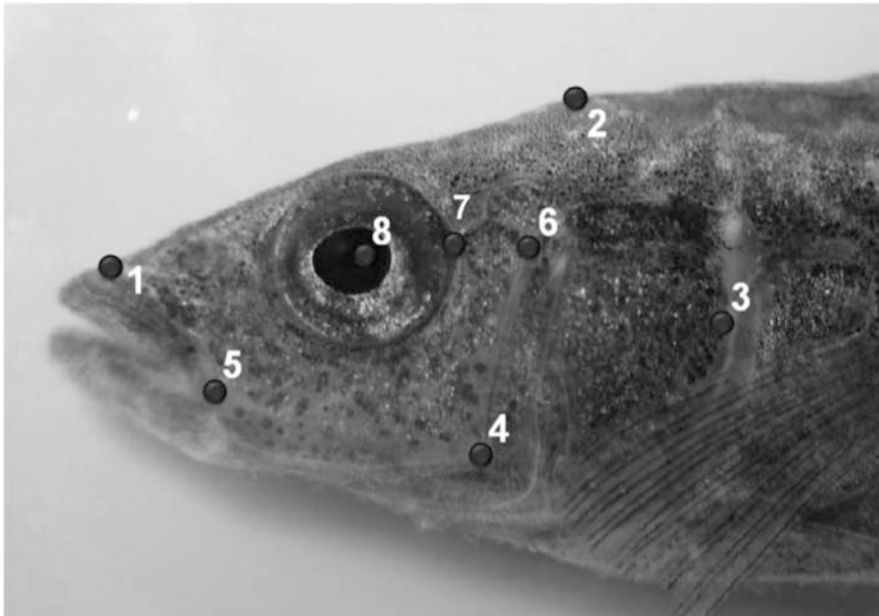
Fig. 10.2 Landmark configuration chosen to characterize head shape among Swiss stickleback populations. 1 = Anterior insertion of premaxilla; 2 = Junction of head & dorsal scales; 3 = Posterior edge of operculum; 4 = Ventral inflexion of preopercular bone; 5 = Posterior extention of premaxilla; 6 = Dorsal extent of preopercular bone; 7 = Anterioventral extent of sphenotic at orbital; 8 = Eye centre

a generalized least-squared Procrustes super-imposition (GPA), which translates all landmark configurations to a common position, resizes them to unit size and corrects them for differences in orientation. Using all corrected landmark configurations together, GPA generates a consensus configuration representing the least squared "average" shape of all specimens (Rohlf and Marcus 1993; Rohlf 1999; Rüber and Adams 2001; Zelditch et al. 2004).

The consensus configuration was then compared to each specimen's corrected landmark configuration using a thin-plate spline. The thin-plate spline calculates interpolation functions or, principal warps, among landmarks that determine the major axes of shape change (Zelditch et al. 2004; Klingenberg 2007). How each specimen's landmark configuration differs from the consensus along these principal warps generates partial warps (i.e., how much movement along each principal warp, in the x and y dimensions, allows corresponding landmarks to touch; Bookstein 1989; Zelditch et al. 2004; Klingenberg 2007). In our case, the thin-plate spline generated 10 partial warps and 2 uniform components representing local (non-affine) and overall (affine) shape changes, respectively (Rohlf 1999; Cavalcanti 2004; Zelditch et al. 2004). Uniform components were found to be significantly different among populations and so retained in further analyses (Wilks $\Lambda = 0.499$, $F_{4,86} = 8.948$ $p < 0.0001$; Cavalcanti 2004; Zelditch et al. 2004; Klingenberg 2007). Significant differences were also observed in the actual sizes of individual fish among the three populations (ANOVA: $F_{2,44} = 10.668$, $p = < 0.0001$); individuals from CLA (lake) were larger than those in both stream populations CUP and SIO (Tukey-HSD, CLA-CUP: $p = < 0.001$, CLA-CUP: $p = 0.002$). SIO and CUP individuals, however, were not statistically different (Tukey-HSD, CUP-SIO: $p = 0.889$). Because size is known to influence shape through ontogenetic stages and allometric trajectories (Klingneberg et al. 1998; Cavalcanti 2004; Zelditch et al. 2004; Klingenberg 2007), shape variables were regressed against size using a multivariate regression and the residuals from this regression were used in subsequent shape analyses (see Klingenberg 2007; Drake and Klingenberg 2008).

A relative warp analysis (PCA on unweighted shape variables; Zelditch et al. 2004; Drake and Klingenberg 2008) was performed on size corrected shape variables to generate a series of relative warps (RWs), which help visualize biologically relatable shape differences among populations (Rohlf et al. 1996; Rohlf 1999; Zelditch et al. 2004; Roy et al. 2007). Deformation of the consensus configuration along each RW was determined by regressing individual shape variables onto RW scores (Klingenberg et al. 2003; Klingenberg 2007; Drake and Klingenberg 2008). A MANOVA was then used on size corrected shape variables to determine differences in head shape among the three populations. Pairwise differences were quantified using Canonical Variates Analyses (CVA) and squared-Mahalanobis distances. CVA rescales overall head shape variation by "within population" variation and rotates the rescaled shape space to align ordination axes with major axes of variation among populations (Klingenberg et al. 2003; Klingenberg 2007). Population-specific deformation grids were also generated using population average scores along the most important RWs (RW1 & RW2) and exaggerated twice

to make visual interpretations easier (Caldecutt and Adams 1998; Cavalcanti 2004; Roy et al. 2007). GM data were collected and analyzed using both the TPS series of programs (tpsUtil, tpsdig2, tpsrelw & tpsregr: Rohlf 2004), Morphoj version 1.00j (Klingenberg 2008), SPSS V16.0 (LEAD technologies, Chicago Ill. USA) and STATISTICA V6.0 (Statsoft, Tulsa. OK, USA).

Feeding Trials

Fish analyzed for shape were also tested for feeding efficiency using both *Tubifex tubifex* (hereafter Tubifex) and various zooplankton (hereafter Zooplankton) representing benthic-stream and pelagic-lake prey items, respectively. Fish were kept under a photoperiod of 11.5:12.5 light:dark and acclimatized to laboratory conditions for 2 weeks prior to feeding experiments. Before feeding trials, fish were placed in trial tanks and fasted for 24 h, habituating them to experimental conditions and instilling hunger. During feeding experiments, individual fish were placed in one of 12 oxygenated 12 L aquaria kept under similar light and temperature regimes as the population holding tanks. Black opaque dividers visually isolated tanks preventing fish from observing one another and minimizing external disturbance.

Feeding experiments were performed in two rounds. In the initial round, half the fish from a given population were tested with Tubifex while the other half received Zooplankton. In the second round food types were swapped. Allocation of fish to an aquarium was random, but food type (i.e., zooplankton or Tubifex) was ramdomly assigned only in the initial round. Subsequent food allocations depended on whether fish were previously offered zooplankton or Tubifex.

Zooplankton Trials

Live zooplankton prey were caught daily from Lake Luzern and size standardized by sieving species > 500 µm but < 1.400 µm. Each day, 5 random zooplankton samples were taken to verify percent composition (data not shown) and 50 individual zooplankton were counted and put into a cold ($\sim 5°C$) water filled Petri dish. Another Petri dish filled with sand was placed at the bottom of the aquarium to standardize the treatments (see Tubifex trials below). Only after the sand was placed at the bottom of the aquaria were the Zooplankton added. The time from zooplankton addition (or clear water in Tubifex treatments) into the aquarium until the first prey item was attacked (Tubifex or Zooplankton) was recorded as the time to first attack (hereafter TTFA). From TTFA, fish were observed for 8 minutes and the number of attacks recorded. Once feeding experiments performed, fish were immediately removed from aquaria and the aquarium water was sieved through a 500 µm mesh. The mesh was washed several times to collect uneaten Zooplankton which were then preserved in ethanol and counted by microscopy.

To estimate recovery efficiency, 50 Zooplankton were added to test aquaria, their contents sieved and the number of Zooplankton recovered was determined. This test was performed 6 times and showed an average 33% Zooplankton recovery per aquaria. Therefore, to conservatively estimate the number of Zooplankton actually eaten by a fish in each trial, estimated number of Zooplankton eaten (*NE*) was adjusted using the following equation:

$$NE = 50 - (Z_r + Z_r(Z_{corr}))$$

where Z_r is the number of Zooplankton items recovered from a trial aquarium and Z_{corr} is the correction factor derived from a general 33% Zooplankton recovery from 50 items according to the following:

$$Z_{corr} = \frac{50 - (50 \times 0.33)}{50 \times 0.33}$$

Tubifex Trials

Twenty Tubifex (obntained from a local pet shop) were placed in sand filled Petri dishes to keep them from floating around the aquarium and simulate benthic feeding conditions. Tubifex were mixed into sand moments before being introduced into the aquarium bottom. As above, additional clear water filled Petri dishes (10 ml at ~5°C) were poured into aquaria standardizing the treatments for all trials. After feeding trials, Petri dishes were removed and the remaining number of Tubifex counted. If a fish did not feed on Tubifex after 30 minutes it was considered to have an efficiency of 0 (see below for feeding efficiency estimates).

Feeding Efficiency Indices

Several feeding efficiency indices were estimated, namely, TTFA (see above), the proportion of prey eaten to that offered (PE), the total number of items eaten (Zooplankton or Tubifex) divided by the number of observed attacks (EPA) and the number of items eaten per attack multiplied by the average number of attacks per minute which determined prey eaten per minute (EPM). All indices were applied to both Zooplankton and Tubifex datasets and are hereafter referred to as ZOOTTFA, ZOOPE, ZOOEPA and ZOOEPM and TUBTTFA, TUBPE, TUBEPA and TUBEPM, respectively. All feeding efficiency data were tested for normality using Kolmogorov-Smirnov tests. Only ZOOEPA required a natural logarithmic transformation and only its transformed data (LNZOOEPA) were used in subsequent analysis.

Analyses

Because size among populations were different (see above), and because size can influence feeding efficiency, we tested whether or not different populations had significantly different feeding efficiencies (TTFA, PE, EPA & EPM) using General Linear Models (GLM) with population, size and their interaction as independent variables. Separate multivariate regressions were also performed regressing size corrected head shape variables with feeding efficiency indices to establish whether or not feeding efficiency was related to head shape differences among populations (Morphoj v 1.00j, Klingenberg 2008; Drake and Klingenberg 2008). Being aware that feeding efficiencies can also be influenced by other more extrinsic factors and may change over the period of the experiment, experimental day, experimental aquaria temperature and their interactions were also used as independent variables to estimate differences in feeding efficiencies among sampled populations.

Defense Traits and Gillrakers

After their use in feeding trials, fish were humanely euthanized and preserved in 90% ETOH. Defense traits (lengths of 1st & 2nd dorsal spines, Pelvic spines, pelvic girdle length, pelvic girdle width and number of lateral plates) and Gillrakers were measured and quantified for all fish from each population ($N = 30$ for CLA, $N = 50$ for CUP and $N = 35$ for SIO). Because of size difference among populations (see above), we applied a size correction to all measured traits except meristic traits which were left as counts. Size correction of traits involved a log transformation and subsequent modification using the following equation:

$$CTV_i = TV_i \left[\frac{GMS}{SL_i} \right]^{AS_{SLvsTV}}$$

where *CTV* is the corrected traits value of an individual (*i*), *TV* is its log transformed trait value, *SL* is its non-transformed standard length, *GMS* is the grand mean of the non-transformed standard length of all individuals from all populations, and *AS* is the average of population specific slopes of the relationship between the trait and standard length. This correction assumes allometric trajectories among populations are similar in direction and slope but is robust to even relatively large deviations and corrects for differences in intercepts (Thorpe 1976; Klingneberg 1998; Hendry et al. 2002). This method is modified from a similar correction derived by Hendry et al. (2002) but adds the principal of normalizing slopes as discussed by Thorpe (1976).

Genetics

Whole genomic DNA was also extracted from the 16 individuals from each population used in feeding trials and from several outgroups populations (Corsica, Danish Marine & Basel, $N = 2$ for each) using the Promega WIZARD DNA Extraction Kit (Promega AG, Dübendorf, Switzerland). Amplification of the cytochrome B (cyt b) and a portion of the control-region (D-loop) were carried out using primers described by Mäkinen et al. (2008). All amplifications followed the same reaction conditions outlined in Mäkinen et al. (2008), but not elaborated here. Resulting PCR products were cleaned using ExoSap protocols (USB Corporation. Staufen, Germany). Sequencing was performed using DTCS Quick start cycle sequencing kit and automated sequencer (CEQ8000) following manufacturer's instructions (Beckman Coulter, Fullerton CA, USA). Generated sequences were aligned using BioEdit version 7.0.9 (Hall 1999) and verified by eye. Phylogenetic reconstruction was performed following a similar protocol as outlined in Mäkinen et al. (2008). Briefly, we used MrBayes 3.1.2 with a random starting tree assuming a general time-reversible (GTR) model of base pair substitution with six substitution rate categories and a variable gamma site distribution as suggested by Ronquist and Huelsenbeck (2003). The analysis was conducted in two parallel runs with 4 MCMC (Markov chain Monte Carlo) chains each, a heating of 0.1 and allowed to run until split frequencies reached a value near 0.01 (after 20 million generations). The final 50% majority rule consensus tree was calculated from remaining trees after a burn in of 10 million generations and support for all nodes was assessed by posterior probabilities.

Results

Head Shape Among Populations

Since we found significant differences in sizes among sampled populations, we tested to see if part of the head shape variation was related to size. Indeed, multivariate regression of shape variables onto size proved to be significant ($p = 0.026$), but only accounted for a small amount (5.25%) of the overall shape variation within and among populations. We therefore used residuals from the size versus shape regression to demonstrate shape differences among populations, and in the remaining analyses.

RW analysis of size corrected head shape variables revealed substantial differences in the shape space occupied by each population (Fig. 10.3a), where SIO had a much larger disparity in shape followed by CUP and then CLA. The shape space outlined by the first two RWs accounted for more than 50% of the overall head shape variation. Shape changes along both RW1 and RW2 were associated with the length and depth of the head and the snout, variation in size and position of the eye and the angle and size of the premaxilla (Fig 10.3a). Differences observed in size corrected
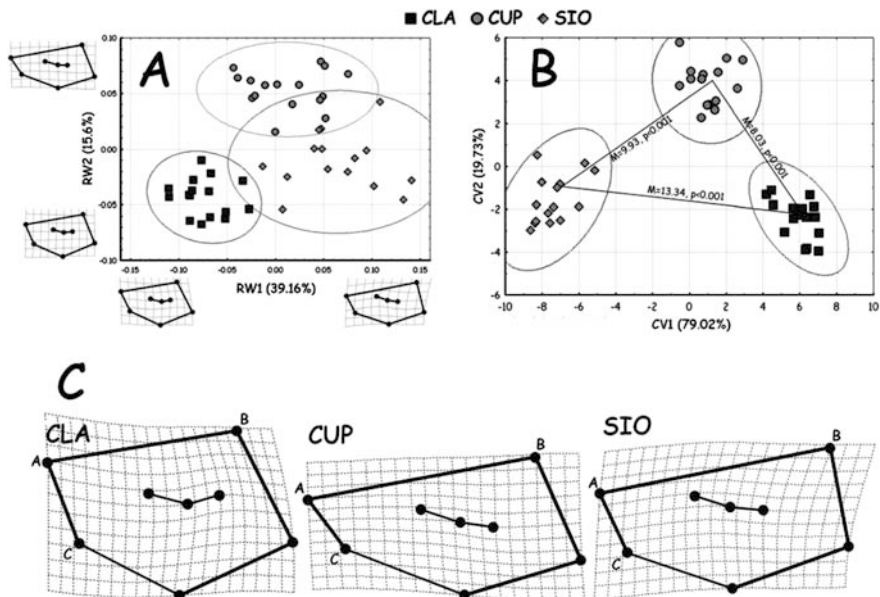
Fig. 10.3 Geometric morphometrics results of head shapes assessed among three Swiss stickle-back populations. (**a**) Relative warps demonstrate that each population occupies distinct areas of shape space and that RW1 & RW2 account for 39.16% and 15.6% of the variation in shape space, respectively. Deformation grids next to axes indicate general shape change trends along each. (**b**) Canonical variates analyses showing statistically significant population separation where CV1 & CV2 account for nearly 100% of among population variation. Lines connecting clusters represent distances between populations (M = Mahalanobis^2 distances, supported by p values derived from 10000 permutations). (**c**) Deformation grids outlining the average head shape deviations of each sampled population from the concensus configuration along RW1 and RW2. Comparing lines AB emphasizes head length differences while AC outlines differences in premaxilla angles and lengths

head shapes among populations were highly significant (Wilk's $\Lambda = 0.003$, $F_{24,68} = 38.56$, $p < 0.0001$). These differences were even more evident using CVA which showed clear separation among population along the first two CVs. CV1 & CV2 accounted for nearly 100% of the variation among populations (Fig. 10.3b) and confirmed the observed RW results showing CUP & SIO expressing more within population variation than does CLA (i.e., larger 95% CI ellipses; Fig. 10.3b). *Posthoc* pairwise comparisons among populations, using squared Mahalanobis distances showed that each population exhibited a significantly different head shape than the others even after sequential Bonferonni correction (Rice 1989) (Fig. 10.3b). Sticklebacks from CLA and CUP were more similar to one another than either was to SIO. A similar distance separating CLA and CUP, separated SIO and CUP, but CUP exhibited a head shape distinct though on the main axis of shape variation always clearly in between, the other two populations. CLA and SIO showed the largest separation in head shape (Fig. 10.3b).

For a more detailed visualization of head shape differences among populations, population-specific deformation grids were generated from population average RW scores along the first two RWs (Roy et al. 2007). Compared with the stream populations CUP and SIO, individuals from the lake population CLA exhibit a shorter, dorso-laterally compressed head (Fig. 10.3c line A–B). The eye is positioned lower within the head and the premaxilla is larger and more upturned than in either the CUP or SIO (Line A–C). In contrast, CUP and SIO populations tend to have more streamlined, elongated heads and eyes set higher within them (Fig 10.3c). These differences are more pronounced in the SIO than in the CUP which is consistent with CUP exhibiting a head shape somewhere between that of CLA and SIO (Fig. 10.3). Although intermediate between the other two populations, CUP does exhibit a distinctly more downturned premaxilla relative to the other two populations (Fig. 10.3c).

Population and Size Effects on Feeding Efficiency

MANOVA used to test for differences in feeding efficiencies among populations and sizes revealed that neither size, nor population or their interaction had any influence on any of the tested feeding efficiency indices (Table 10.1).

Based on these results, size was considered to have a negligible effect on feeding efficiencies among the sampled stickleback populations and was thus excluded from further analyses (other than its correction applied to shape, see methods).

Head Shape and Feeding Efficiencies

Multivariate regressions of size corrected head shape variables onto feeding efficiencies demonstrated that head shape was not related to any of the feeding efficiency indices regardless of treatment (i.e., Tubifex or Zooplankton). Thus, head shape differences among populations determined using geometric morphometrics, were not statistically related to feeding efficiencies (Fig. 10.4).

As shown by the deformation grids, however, the position of the eye differs markedly among the three populations (Fig. 10.3c). It may be that although head shape differences among populations were not related to handling and consumption of the prey, they may nevertheless be related to the detection or recognition of food items (Walton et al. 1994). To test this, shape variables were regressed onto the TTFA, our proxy of prey detection and recognition. TTFA, however, was not significantly related to head shape for either Tubifex or Zooplankton treatments (Fig. 10.4; $P > 0.05$). TTFA was also tested against feeding efficiency indices to see if the detection of prey items influenced these latter. No such relationship was significant, however, indicating that TTTA was not a factor driving efficiencies as assessed in our experiments (TUBPE- $R^2 = 0.001$, $F_{1,47} = 0.065$, $p = 0.799$; TUBEPA- $R^2 = 0.004$, $F_{1,47} = 0.187$, $p = 0.668$; TUBEPM- $R^2 = 0.001$, $F_{1,47} = 0.065$, $P = 0.799$; ZOOPE- $R^2 = 0.003$, $F_{1,47} = 0.137$, $p = 0.713$; LNZOOEPA- $R^2 = 0.004$, $F_{1,47} = 0.191$, $p = 0.664$; ZOOEPM- $R^2 = 0.003$ $F_{1,44} = 0.137$, $p = 0.713$).

Table 10.1 GLM results of population, size and their interaction on the various estimated feeding efficiencies for both *Tubifex tubifex* (TUB) and zooplankton (ZOO) trials

Factors	TUBPE		TUBEPA		TUBEPM		ZOOPE		ZOOEPA		ZOOEPM	
	F	p	F	p	F	p	F	p	F	p	F	p
Population	0.415	0.663	2.086	0.137	0.421	0.659	0.076	0.927	0.182	0.835	0.210	0.812
Size	1.991	0.166	0.858	0.360	1.692	0.200	0.023	0.879	1.415	0.241	0.040	0.843
Population × Size	0.430	0.654	1.888	0.164	0.440	0.647	0.131	0.877	0.159	0.853	0.229	0.796
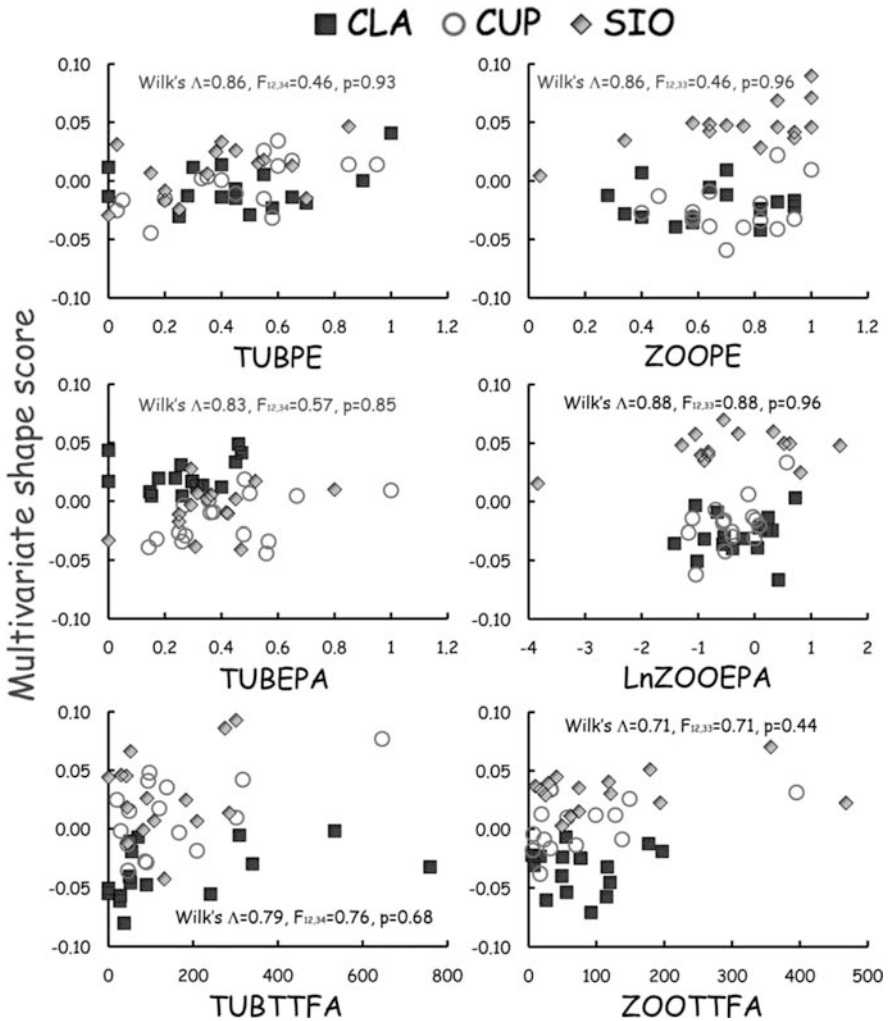
Fig. 10.4 Multivariate shape scores (calculated as per Drake and Klingenberg 2008) outlining the relationship between shape and various feeding efficiency measures estimated among sampled Swiss stickleback populations. Trials performed with typical stream-benthic prey (*Tubifex tubifex*) and lake-pelagic prey (zooplankton, various species). None of the realtionship tested proved to be significant. TUBEPM & ZOOEPM showed patterns nearly identical to those of TUBPE & ZOOPE, respectively (TUBEPM *Wilk's* $\Lambda = 0.88$, $F_{12,34} = 0.38$, $p = 0.96$, ZOOEPM *Wilk's* $\Lambda = 0.67$, $F_{12,33} = 1.38$, $p = 0.22$)

No significant relationships were detected when relating different feeding effi-ciencies in either treatment to extrinsic factors such as water temperature in test aquaria and the days over which the experiments were run, nor when considering the interaction among them (Table 10.2). Hence, extrinsic factors were unlikely confounding our tests. Thus, although significant differences in head shapes were

 observed among the sampled Swiss stickleback populations, these differences were not related to either the ability to detect prey, or to the feeding efficiency on the types of prey offered.

Defensive Traits

MANOVA run on the defensive traits and gill raker counts measured in each population proved to be unreliable (*Box* $M_{56,26661} = 123.36$, $p < 0.0001$). Therefore, univariate ANOVAs were used to tests for differences among populations for each trait. Significant differences were observed for all defensive traits, except for pelvic girdle width ($F_{2,44} = 0.862$, $p = 0.85$; Fig. 10.5). Gill raker counts were also significantly different ($F_{2,44} = 4.5$, $p = 0.013$), where CLA was different from SIO but CUP was not different from either (Fig. 10.5). In all cases of spine lengths, CLA exhibited significantly longer spines than either CUP or SIO, but differences between CUP and SIO were negligible and not significant (Fig. 10.5). The pelvic girdle length in CLA was also significantly longer than in CUP, which in turn was longer than that in SIO (Fig. 10.5). Lateral plate number, however, was not significantly different in either population from the Constance area, but both had significantly greater numbers of plates than did the individuals from SIO (Fig. 10.5). Generally, all defence-related traits were higher in the lake population CLA then in either stream population.
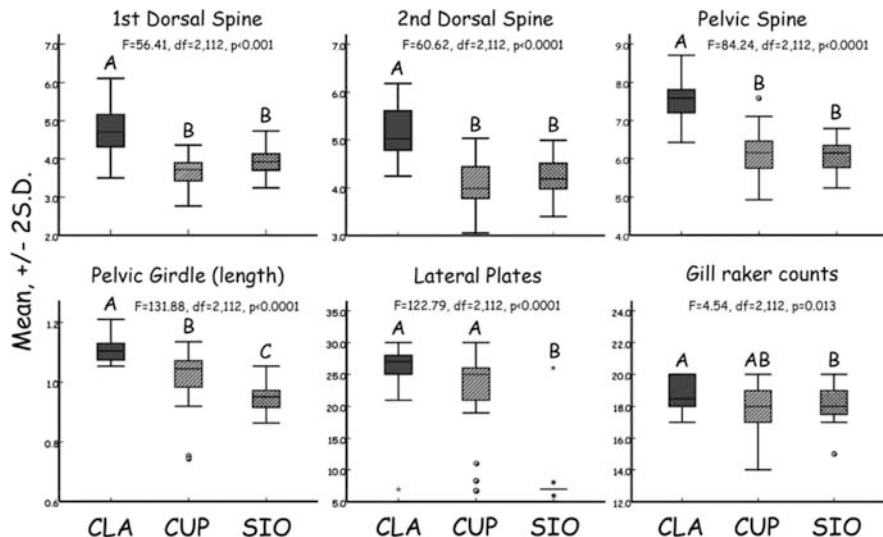
Fig. 10.5 Comparison among size corrected defensive traits measured among three sampled stickleback populations. *F*-values from univariate two-tailed ANOVA reported for each trait. PGW (pelvic girdle width) – data not shown ($F = 0.86$, $df = 2$, $p = 0.085$). *Post-hoc* Tukey's HSD performed on all traits, and populations with different superscript letters indicate significant differences in trait

Table 10.2 GLM results of experimental day (Exp Day), experimental tank temperature (Exp Temp) and their interaction (Day × Temp) on the various estimated feeding efficiencies for both *Tubifex tubifex* (TUB) and zooplankton (ZOO) trials

Factors	TUBPE		TUBEPA		TUBEPM		ZOOPE		ZOOEPA		ZOOEPM	
	F	p	F	p	F	p	F	p	F	p	F	P
Exp Day	0.379	0.907	0.976	0.465	1.048	0.418	0.691	0.679	0.614	0.74	0.691	0.679
Exp Temp	0.041	0.841	2.267	0.142	0.004	0.948	0.827	0.369	0.283	0.598	0.827	0.370
Day × Temp	0.384	0.905	0.961	0.475	1.038	0.424	0.747	0.635	0.623	0.733	0.747	0.635

Genetics

The phylogenetic tree based on Bayesian inferences showed that all CLA and CUP individuals share the same haplotype and that this haplotype is more closely related to the Danish (Marine) sticklebacks than to those sampled from the purported progenitive Basel population (Fig. 10.6).

Although more haplotypes are observed in the SIO population, all appear to be similarly distant from the Basel haplotype but substantially different from that observed in the Constance area (Fig. 10.6). Thus results indicate both the population SIO and those from the Constance area belong to two distinct lineages which are both more closely related to geographically distant haplotypes (Danish Marines & Corsican) than they are to the purported initial Swiss colonizer.
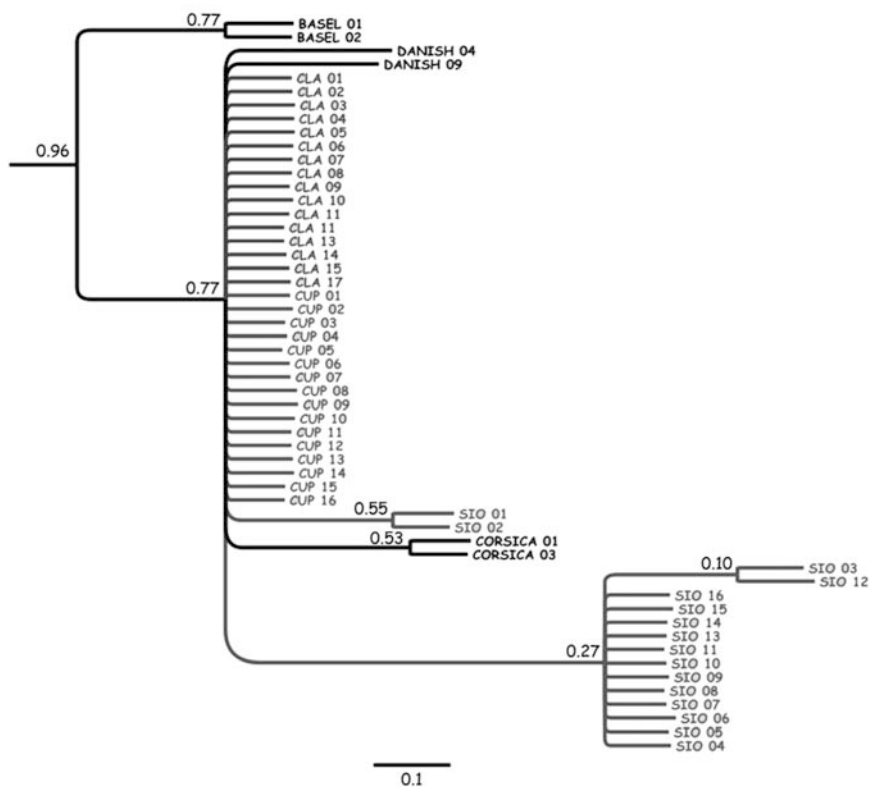
Fig. 10.6 50% majority rule concensus tree of three Swiss stickleback populations including 3 additional outgroups based on a Bayesian inference phylogeny. Posterior probabilities indicated above the nodes (DANISH = Danish Marine population, CORSICA = Mediteranean freshwater population and BASEL = Original native Swiss population). Relationship derived from mtDNA sequences of D-loop & cytochrome B. Population lineages (CLA, CUP & SOI) show closer relationships with more distant outgroups (DANISH & CORSICA) than with Basel population and thus likely originate from different colonizing lineages

Discussion

When divergent selection drives populations to exploit different food resources, differences in overall body morphology or specific trophic morphology often emerge (Bentzen and MacPhail 1984; Witte 1984; Schluter 1993, 1995; Robinson 2000; Danley and Kocher 2001). If divergent natural selection filters variation in morphological traits providing efficient feeding on alternative food resources (e.g., lake versus stream or benthic versus limnetic), a correlation between feeding morphology and feeding efficiency is predicted (Bentzen and McPhail 1984; Schluter 1993, 2000; Bernatchez 2004; Coyne and Orr 2004). Here, we focused on differences in head shape since feeding performance and head shape are likely tightly linked (Bentzen and McPhail 1984; Hart and Gill 1994; Schluter 1993, 1995; Robinson 2000; Rincòn et al. 2007). We show that head shape differed significantly among the three sampled Swiss stickleback populations and that this was only partially associated with differences in overall body size. Yet, no significant differences were found among sampled populations in their feeding efficiencies on either limnetic or benthic prey, nor did we find any significant relationship linking head shape variation to variation in feeding efficiency (Table 10.1, Fig. 10.4).

Head shapes were also tested for a relationship with prey detection, because deformation grids indicated marked differences among populations in their relative eye positions (Fig. 10.3c). We postulated that variation in head shape may be associated with variation in detecting different prey (Walton et al. 1994; Walker 1997; Caldecutt and Adams 1998; Walker and Bell 2000; Roy et al. 2007). Time to first attack (TTFA) was used as a surrogate for prey detection, but again, no relationship was found with head shape, nor did TTFA have any influence on other feeding efficiency estimates. These results are not consistent with the hypothesis that head shape divergence among these populations was driven by ecological selection for efficient feeding on limnetic/lake versus benthic/stream food types (Schluter 2000; Bernatchez 2004). Thus, local adaptation of the three populations to their respective feeding environments does not appear to be the cause of the expressed head shape differences in this case.

This lack of relationship between head shape and feeding efficiency is comparable to results from Day and McPhail (1996) who observed only weak association between morphological specificity and foraging efficiency and fitness in other stickleback populations reared under laboratory settings. These authors suggested that while morphological changes are more difficult to achieve, behavioural plasticity can occur quite rapidly in most populations and is easily reversible when population are confronted with sudden diet shifts such as those imposed in our study. The lack of feeding efficiency differences among Swiss populations could be explained by their ability to easily adjust feeding behaviour to the prey presented in experiments, regardless of head shape. Hatfield and Schluter (1999) also demonstrated weak to no significant fitness advantages of benthic or limnetic sticklebacks, or their hybrids, when raised in simulated benthic or limnetic environments in a laboratory setting. However, Hatfield and Schluter (1999) and Rundle (2002), showed marked differences in limnetic, benthic, F1 hybrids and backcrosses in fitness estimates when

tested under more natural conditions. Hendry et al. (2002) demonstrated similar findings when testing differences among open-lake, lake-littoral and stream populations in reciprocal transplant experiments run in situ using enclosures. Results from this and the above cited studies suggest that conditions causing significant performance tradeoffs among stickleback ecotypes are absent, or at the very least, difficult to reproduce under laboratory conditions and that more natural environments tend to better manifest differences in feeding efficiencies and resulting fitness differences among ecotypes (Schluter 1993, 1995; Day and McPhail 1996; Hatfield and Schluter 1999; Hendry et al. 2002; Rundle 2002). Under more natural conditions, factors such as parasite loads, inter and intra-specific competition and predation pressure, along with other extrinsic factors, act synergistically to substantially alter feeding performance. Natural conditions "sharpen" the adaptive peaks of the selection landscape between poorly and better adapted ecotypes within a given environment making efficiency differentials more obvious (Vamosi 2002; Rundle et al. 2003). Thus, although the sampled populations exhibited differences in head morphology which may be related to increased feeding efficiencies within population-specific environments, this association may not have been observed because laboratory conditions did not sufficiently simulate natural gradients where sampled populations experience environment-specific trade-offs.

However, even if our experimental design was not sufficient to simulate adequate conditions to observe efficiency differences related to head shape, we would nevertheless expect to observe head shape patterns among populations resembling those observed in other studies quantifying stickleback feeding morphology among lake and stream ecotypes, including those using GM (Walker 1997; Caldecutt and Adams 1998; Walker and Bell 2000; Berner et al. 2008). Although we showed that for some head features, sampled populations do indeed adhere to the repeated pattern, for other features they do not. The eye and mouth positions in CLA versus CUP and SIO are consistent with other studies reporting more upturned mouths and more sunken eyes in lake versus stream resident populations. However, the overall deeper and shorter head and snout of CLA versus CUP and SIO are opposite to previously reported trends (Walker 1997; Caldecutt and Adams 1998; Walker and Bell 2000). Moreover, if local adaptation to respective feeding environments was indeed a dominant force selecting for different head features we would expect both stream populations (CUP & SIO) to be convergent. Although significant differences in head shapes between CUP and SIO, were observed, including a more extreme head elongation and a smaller more elevated eye in SIO, and a much more downturned mouth in CUP, the two stream populations indeed showed trends of convergence when compared to the lake population, despite the closer phylogenetic relatedness of CUP to the nearby lake population CLA (Fig. 10.3).

Although substantial differences are clear between both stream populations, they nevertheless follow a general trend away from the lake ecotype shape with many features in common. The respective magnitudes and trajectories of divergence for both stream populations in relation to CLA are somewhat different, but this is not necessarily unexpected considering recent evidence demonstrating variable specific direction and magnitude of divergence among other replicated lake and stream

paired stickleback populations (Hendry et al. 2002; Hendry and Taylor 2004; Berner et al. 2008). Hendry and Taylor (2004) and Berner et al. (2008) show that although the general trends in phenotypic divergence between paired stream and lake resident populations in several different watersheds (up to 8) from Vancouver Island (BC, Canada) are consistent, the specifics of each pair can be quite distinct. Thus, the differences between CUP and SIO versus CLA may be explained by the fact that these former populations have different versus common recent ancestry (see below) relative to CLA.

Another reason for the lack of relationships between head shape and feeding efficiency in the sampled populations may be that variation in food availability is not the primary source of divergent selection between these populations. Stickleback populations are more than likely diverging along more than a single ecological axis (Rundle and Nosil 2005; Nosil et al. 2009). Previous work has demonstrated that stickleback populations often differ in morphology associated with different predation regimes and that different defense morphologies mediate tradeoffs between protection against vertebrate versus invertebrate predators, and perhaps tradeoffs between protection and foraging efficiency (Vamosi 2002; Rundle et al. 2003; Nosil and Reimchen 2005; Barrett et al. 2008; Marchenko et al. 2009). Large numbers of bony plates and longer spines are typically associated with marine forms of sticklebacks and with freshwater populations exposed to high levels of vertebrate predation (birds and fish; Reimchen 1994; Vamosi 2002; Rundle et al. 2003; Nosil and Reimchen 2005). Long spines and many bony plates make prey handling difficult for gape-limited predators. However, more bony plates and longer spines are costly to produce, provide effective anchors for insect predators (which dominate in many freshwater environments) and make maneuverability among macrophytes and benthic litter refugia more difficult (Reimchen 1994; Vamosi 2002; Rundle et al. 2003; Nosil and Reimchen 2005; Marchenko et al. 2009). Here, we showed that the three populations sampled differed significantly in five of the six quantified defensive traits (Fig. 10.5). Moreover, in all instances of spine lengths, both stream resident populations were convergent and had smaller spines than those of the lake resident. Although not all pelvic girdle traits were different among populations, those that were followed similar trends as the spine lengths (Fig 10.5). Taken together, these results indicate that the morphological divergence among populations may in fact be driven partly by adaptation to predation, or by the interaction of factors including both predation and feeding resources. This is consistent with a multifaceted selection type regime (Rundle et al. 2003; Nosil et al. 2009).

Although defensive traits have been correlated to survival success and habitat specific predation regimes, the efficiency of specific defensive traits as predator deterrents is difficult to assess (i.e., see Vamosi 2002; Rundle et al. 2003; Nosil and Reimchen 2005; Barrett et al. 2008; Marchenko et al. 2009). Future research is needed to determine how head and overall body shapes, as well as defensive traits, are related to feeding efficiency under different predation regimes within the lake-stream transition. Such an experiment could at once confirm the utility of individual traits both separately and in combination, and effectively demonstrate the synergistic influence of various selection pressures in lake – stream population divergence

(Rundle et al. 2003; Nosil et al. 2009). Again, however, just as in the feeding traits, some defense traits support in-situ divergent local adaptation of lake and stream eco-types during the Swiss stickleback invasion, while others do not. Lateral plates, in particular, are divergent between Constance and Geneva area populations rather than between stream and lake ecotypes (Fig 10.5). This trait may in fact reflect the differ-ent ancestry of Constance and Geneva populations as suggested from the mtDNA phylogeny (Fig. 10.6). Thus, some evidence points to a degree of local adaptation in head shapes and defensive traits, but both trait types also seem to carry a significant signature of historical contingency.

Switzerland was recently colonized by sticklebacks from multiple sources representing different European lineages that were probably adapted to very different native environments (Fig. 10.6). Given that sticklebacks have only been for approximately 100–150 generations in the Swiss systems we studied here (Fatio 1882; Bertin 1927), it is not surprising that differences in traits among populations have a large historical component that depends on the founding populations (Coyne and Orr 2004; Seehausen 2007). This explanation could certainly account for differ-ences observed between SIO and both populations from the Constance area (CLA & CUP), which have been seeded by very different lineages. However, contingency does not explain the observed differences between CUP and CLA. Absence of divergent local adaptation within lineages that split only 100–150 generations ago would predict strong similarities between CUP and CLA. Yet, significant head shape and defensive trait differences were observed between these populations suggesting some degree of divergence between them despite the short time that was available. This can only be explained by divergent selection, the required strength of which depends on the extent of gene flow (Hendry et al. 2002, 2004; Seehausen et al. 2008; Nosil et al. 2009). Estimating how much gene flow exists and how much this constrains adaptive divergence between populations from common colonizing lin-eages could resolve this issue (Hendry et al. 2002; Hendry and Taylor 2004; Moore et al. 2007). Future studies should focus on gene flow among adjacent populations following the lake stream gradient in many independently colonized lineages. Such investigations could clarify the relative magnitudes of different sources of divergent selection, their order of influence, and the constraining or mediating effects of gene flow on the evolutionary response to these selection pressures (Schluter 2000; Hendry et al. 2004; Seehausen 2007; Nosil et al 2009). These studies could be easily performed in Swiss aquatic ecosystems wherein many have only recently been colonized either by the same, by relatively distantly related stickleback lineages or by both that are coming into secondary contact (Lucek 2009).

Conclusion

This study provides a test of the expected relationship that should exist between feeding morphology (described here as head shape) and feeding efficiency within and between populations occupying ecologically different environments. No signif-icant relationship was observed in the tested Swiss stickleback populations, likely

related to a variety of reasons including possible insensitivity of the experimental design. The most likely reason, however, stems from the interaction of historical contingency and contemporaneous adaptation in explaining the phenotypic variation between populations in the recent invasion of Swiss systems by sticklebacks. Thus, we found that head shapes of Swiss sticklebacks are indeed highly divergent between populations, but the underlying causes of this variation remain ambiguous and may partly be the result of adaptation to their native environments outside the current Swiss distribution, and partly of ecological selection acting on more morphological features than simply those related to food resource use. We show that geometric morphometrics is an invaluable technique to investigate fundamental principles in the adaptive radiation framework.

References

Adams DC, Rohlf FJ & Slice DE (2004) Geometric morphometrics: ten years of progress following the 'revolution'. Italian Journal of Zoology. 71: 5–16.

Albertson RC, Streelman JT & Kocher TD (2003) Genetic basis of adaptive shape differences in the cichlid head. Journal of Heredity. 94: 291–301.

Bakker TCM (1993) Choosy female sticklebacks generate a positive genetic correlation between preference and preferred ornament. Nature. 363: 255–257.

Barret RDH, Rogers SM & Schluter D (2008) Natural selection on a major armor gene in threespine stickleback. Science. 322: 255–257.

Bell MA & Foster SA (1994) The evolutionary biology of the threespine stickleback, Oxford University Press, Oxford, MA.

Bentzen P & McPhail JD (1984) Ecology and evolution of sympatric sticklebacks (Gasterosteus): specialization for alternative trophic niches in the Enos Lake species pair. Canadian Journal of Zoology. 62: 2280–2286.

Bernatchez L (2004) Ecological theory of adaptive radiation. An empirical assessment from Coregonine fishes (Salmoniformes). In Evolution Illuminated, Salmon and Their Relatives. (Hendry AP & Sterns S) p75–207. Oxford University Press, New York.

Berner D, Adams DC, Grandchamp AC & Hendry AP (2008) natural selection drives patterns of lake-stream divergence in stickleback foraging morphology. Journal of Evolutionary Biology. 21: 1653–1665.

Bertin L (1927) Recherches Bionomiques, biométriques et systématiques sur les épinoches (*Gastéréidés*). Annales de L'Institut Océanographique. La Sorbonne Paris France.

Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformation. IEEE Transactions on Pattern Analysis & Machine Intelligence. 11: 567–585.

Caldecutt WJ & Adams DC (1998) Morphometrics of trophic osteology in the threespine stickleback, *Gasterosteus aculeatus*. Copeia. 1998: 827–838.

Cavalcanti MJ (2004). Geometric morphometric analysis of head shape variation in four species of hammerhead sharks (Carcharhiniformes:*Sphyrnidae*). In Morphometrics – Applications in Biology and Paleontology. (Elewa AMT) p97–113. Springer Verlag, Heidelberg.

Coyne JA & Orr HA (2004) Speciation. Sinauer Associates, Inc., Sunderland, MA.

Danley PD & Kocher TD (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi. Molecular Ecology. 10: 1075–1086.

Drake AG & Klingenberg CP (2008) The pace of morphological change: historical trasformation of skull shape in St Bernard dogs. Proceedings of the Royal Society Biological Sciences. 275: 71–76.

Day T & McPhail JD (1996) The effect of behavioural and morphological plasticity on foraging efficiency in the threespine stickleback (*Gasterosteus* sp). Oecologia. 108: 380–388.

Fatio V (1882) Faune des vertébrés de la Suisse. H. Georg, Geneva, Switzerland.

Hall T (1999) BioEdit v7.0.9. Ibis Bioscience., Carlsbad, CA.

Hart PJB & Gill AB (1994) Evolution of foraging behaviour in the threespine stickleback. In The Evolutionary Biology of the Threespine Stickleback (Bell MA & Foster SA) p207–239. Oxford University Press, New York.

Hatfield T & Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution. 53: 866–873.

Hendry AP & Taylor EB (2004) How much of the variation in adaptive divergence can be explained by gene flow? An evaluation using lake-stream stickleback pairs. Evolution 58: 2319–2331.

Hendry AP, Taylor EB & McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the misty system. Evolution. 56: 1199–1216.

Hudson AG, Vonlanthen P, Müller R & Seehausen O (2007) Review: The geography of speciation and adaptive radiation in coregonines. Advances in limnology. 60: 111–146.

Kitano J, Bolnick DI, Beauchamp DA, Mazur MM, Mori S, Nakano T & Peichel CL (2008) Reverse evolution of armor plates in the threespine stickleback. Current Biology. 18: 769–774.

Klingenberg CP (2008) Morphoj version 1.00j. Java version 1.5.0_16 for Mac OS X 10.5.6 University of Manchester, Manchester, UK.

Klingenberg CP (2007) Analysis of organismal form: an introduction to morphometrics, web-based course notes, University of Manchester, Manchester, UK.

Klingenberg CP, Barluenga M & Meyer A (2003) Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. Biological Journal of the Linnean Society. 80: 397–408.

Klingneberg CP (1998) Multivariate allometry. In Advances in Morphometrics. (Marcus LF, Corti M, Loy A, Naylor, GJP & Slice DE) p23–49. Plenum Press, NewYork.

Kraak SBM, Mundwiler B & Hart PJB (2001) Increased number of hybrids between benthic and limnetic three-spined sticklebacks in Enos Lake, Canada; the collapse of a species pair? Journal of Fish Biology. 58: 1458–1464.

Lucek KOL (2009) Genetic History and Phenotypic Diversity of a Recent Invasion: The Three-spined Stickleback in Switzerland. M.Sc. Thesis, University of Bern, Bern Switzerland.

Mäkinen HS & Merilà J (2008) Mitochondrial DNA phylogeography of the three-spined stickleback (*Gasterosteus aculeatus*) in Europe—Evidence for multiple glacial refugia. Molecular Phylogenetics and Evolution. 46: 167–182.

Mäkinen HS, Cano JM & Merilà J (2006) Genetic relationships among marine and freshwater populations of the European three-spined stickleback (*Gasterosteus aculeatus*) revealed by microsatellites. Molecular Ecology. 15: 1519–1534.

Marchenko KB (2009) Predation's role in repeated phenotype and genetic divergence of armor in three-spined stickleback. Evolution. 63: 127–138.

McKinnon JS & Rundle HD (2002) Speciation in nature: the threespine stickleback model systems. Trends in Ecology & Evolution. 17: 480–488.

Milinski M & Bakker TCM (1990) Female sticklebacks use male colouration in mate choice and hence avoid parasitized males. Nature. 344: 330–333.

Moore JS, Gow JL, Taylor EB & Hendry AP (2007) Quantifying the constraining influence of gen flow on adaptive divergence in the lake-stream threespine stickleback system. Evolution. 61: 2015–2026.

Munzig J (1963) Evolution of variation and distributional patterns in european populations of the tree-spined stickleback, *Gasterosteus aculeatus*. Evolution. 17: 320–332.

Nosil P, Harmon LJ & Seehausen O (2009) Ecological explanations for (incomplete) speciation. Trends in Ecology & Evolution. 24: 145–156.

Nosil P & Reimchen TE (2005) Ecological opportunity and levels of morphological variance within freshwater stickleback populations. Biological Journal of the Linnean Society. 86: 297–308.

Peichel CL, Nereng KS, Ohgi KA, Cole BLE, Colosimo PF, Buerkle CA, Schluter D & Kingsley DM (2001) The genetic architecture of divergence between threespine stickleback species. Nature. 414: 901–905.

Reimchen TE (1994) Predators and morphological evolution in threespine stickleback. In The Evolutionary Biology of the Threespine Stickleback (Bell MA & Foster SA) p240–276. Oxford University Press, New York.

Reusch TBH, Wegner KM & Kalbe M (2001) Rapid genetic divergence in postglacial populations of threespine stickleback (*Gasterosteus aculeatus*): the role of habitat type, drainage and geographical proximity. Molecular Ecology. 10: 2435–2445.

Rice WR (1989) Analyzing tables of statistical tests. Evolution. 43: 223–225.

Rincòn PA, Bastir M & Grossman GD (2007) Form and performance: body shape and prey-capture success in four drift-feeding minnows. Oecologia. 152: 345–355.

Robinson BW (2000) Trade offs in habitat-specific foraging efficiency and the nascent adaptive divergence of sticklebacks in lakes. Behaviour. 137: 865–888.

Rohlf FJ (2004) TPS program series available at http://life.bio.sunysb.edu/morph/

Rohlf FJ (1999) Shape statistics: procrustes superimpositions and tangent spaces. Journal of Classification. 16: 197–223.

Rohlf FJ, Loy A & Corti M (1996) Morphometric analysis of old world talpidae (Mammalia, Insectivora) using partial-warp scores. Systematic Biology. 45: 344–362.

Rohlf FJ & Marcus LF (1993) A revolution in morphometrics. Trends in Ecology & Evolution. 8: 129–132.

Rohlf FJ (1990) Morphometrics. Annual Review of Ecology and Systematics. 21: 299–316.

Ronquist F & Huelsenbeck JP (2003) MrBayes 3: inference under mixed models. Bioinformatics. 19: 1572–1574.

Roy D, Paterson G, Hamilton PB, Heath DD & Haffner GD (2007) Resource-based adaptive divergence in the freshwater fish *Telmatherina* from Lake Matano, Indonesia. Molecular Ecology. 16: 35–48.

Rüber L & Adams DC (2001) Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. Journal of Evolutionary Biology. 14: 325–332.

Rundle HD, Vamosi SM & Schluter D (2003) Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. Preceedings of the National Academy of Sciences USA. 100: 14943–14948.

Rundle HD (2002) A test of ecologically dependent postmating isolation between sympatric sticklebacks. Evolution. 56: 322–329.

Rundle HD & Nosil P (2005) Ecological speciation. Ecology Letters. 8: 336–352.

Seehausen O (2007) Chance, historical contingency and ecological determinism jointly determine the rate of adaptive radiation. Heredity. 99: 361–363.

Seehausen O, Takimoto G, Roy D & Jukela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. Molecular Ecology. 17: 30–44.

Schluter D (2000) The ecology of adaptive radiation. Oxford Series in Ecology and Evolution, New York.

Schluter D (1995) Adaptive radiation in sticklebacks – trade-offs in feeding performance and growth. Ecology. 76: 82–90.

Schluter D & Nagel LM (1995) Parallel speciation by natural selection. The American Naturalist. 146: 292–301.

Schluter D (1994) Experimental-evidence that competition promotes divergence in adaptive radiation. Science. 266: 798–801.

Schluter D (1993) Adaptive radiation in sticklebacks – size, shape, and habitat use efficiency. Ecology. 74: 699–709.

Skùlason S, Snorrason SS, Ota D & Noakes DLG (1993) Genetically based differences in foraging behavior among sympatric morphs of arctic charr (Pisces, Salmonidae). Animal Behaviour. 45: 1179–1192.

Taylor EB & McPhail JD (2000) Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. Proceedings of the Royal Society of London Series B-Biological Sciences. 267: 2375–2384.

Thorpe RS (1976) Biometric analysis of geographic variation and racial affinities. Biological Reviews. 51: 407–452.

Vamosi SM (2002) Predation sharpens the adaptive peaks: survival trade-offs in sympatric sticklebacks. Annales Zoologici Fennici. 39: 237–248.

Vonlanthen P, Roy D, Hudson AG, Largiader CR, Bittner D & Seehause O (2009) Divergence along a steep ecological gradient in lake whitefish (coregonus sp.). Journal of Evolutionary Biology. 22: 498–514.

Walker JA & Bell MA (2000) Net evolutionary trajectories of body shape evolution within a microgeographic radiation of threespine stickleback (*Gastersteus aculeatus*). Journal of the Zoological Society of London. 252: 292–302.

Walker JA (1997) Ecological morphology of lacustrine treespine stickleback Gasterosteus aculeatus L. (Gasterosteidae) body shape. Biological Journal of the Linnean Society. 61: 3–50.

Walton WE, Easter Jr SS, Malinowski C & Hairston Jr NG (1994) Size-related change in the visual resolution of sunfish (Lepomis spp.). Canadian Journal of Fisheries & Aquatic Sciences. 51: 2017–2026.

Witte F (1984) Ecological differentiation in Lake Victoria haplochromines: comparison of cichlid species flocks in African Lakes. In Evolution of Fish Species Flocks. (Echelle AA & Kornfield I) p155–168. University of Maine Press, Orono.

Zelditch ML, Swiderski DL, Sheets HD & Fink WL (2004) Geometric morphometrics for biologists: a primer. Elsevier, Academic Press, Amsterdam.

Chapter 11
Macroevolutionary Trends in the Skull of Sauropodomorph Dinosaurs – The Largest Terrestrial Animals to Have Ever Lived

Mark T. Young and Matthew D. Larvan

Idea and Aims

Geometric morphometric analysis was applied to the skull (separately in lateral and dorsal view) of sauropodomorph dinosaurs from the Mesozoic. These analyses were applied to quantify the macroevolutionary trends in sauropodomorph craniofacial form. As the early craniofacial evolution of Sauropodomorpha has been considered to be "conservative", with all major changes to skull shape occurring within Sauropoda itself, geometric techniques are herein employed to quantify this statement. The landmark-based approach (relative warps analysis) was able to confirm this statement, albeit within the limits of the sample of this study. The morphospaces in both views demonstrate that sauropodomorphs became more divergent in terms of craniofacial form, with the sauropods occupying a distinct region of, and a greater volume of, morphospace than the "prosauropods". Although the concordance between craniofacial form and phylogeny is statistically significant, the correlations themselves are not significant. As such, craniofacial evolution within Sauropodomorpha is "shaped" by more than just historical contingency.

Introduction

Throughout most of the Mesozoic era (250–65 Mya) sauropodomorph dinosaurs were the dominant terrestrial herbivores. They reached the biomechanical limits of terrestrial gigantism, with many species of taxa achieving sizes an order of magnitude larger than other extinct or living terrestrial animals (Sander 2000). Therefore, the question arises, how did these multi-tonne giants fuel their colossal bodies? How was it possible for there to be six or more contemporaneous species of giant

M.T. Young (✉)
Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queen's Road, Bristol, BS8 1RJ, UK; Department of Palaeontology, Natural History Museum, Cromwell Road, London, SW7 5BD, UK
e-mail: mark.young@bristol.ac.uk

sauropods in Late Jurassic (155–145 Mya) ecosystems? As with examples of multiple contemporaneous large carnivorous dinosaurs (e.g. Henderson 1998), resource (niche) partitioning via morphological differentiation has been posited as maintaining high biodiversity. In the case of sauropods, differentiation is concentrated on craniofacial form (although tooth crown and neck morphology are also important) (e.g. Barrett and Upchurch 1994; Calvo 1994; Upchurch and Barrett 2000; Barrett and Upchurch 2007; Upchurch et al. 2007).

Sauropodomorpha is a clade consisting of two well-known groups, Sauropoda and Prosauropoda. Recent phylogenetic analyses have convincingly demonstrating the paraphyly of Prosauropoda (e.g. Upchurch et al. 2007; Yates 2003, 2007; Yates and Kitching 2003), therefore we hereafter refer to all non-sauropod sauropodomorphs as "prosauropods" in the knowledge they do not constitute a natural group. Although the topology of the sauropodomorph evolutionary tree is still disputed in these analyses, the taxa selected in this study remain stable throughout (Fig. 11.1).

Although recent work has demonstrated the unrealised craniofacial diversity within Sauropoda (e.g. *Nigersaurus*; Sereno et al. 2007), "prosauropods" have long been considered conservative in cranial morphology, while their feeding mechanisms are still considered to be conservative (e.g. Upchurch and Barrett 2005). As
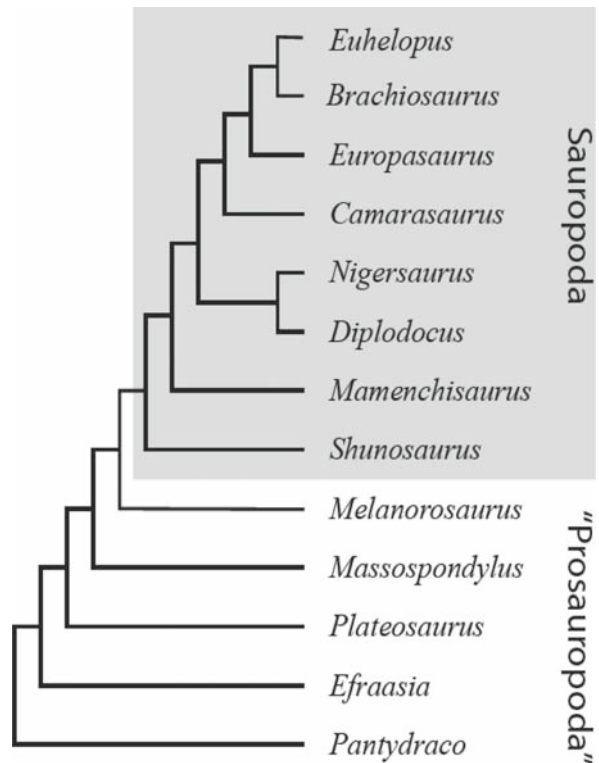
Fig. 11.1 Evolutionary relationships of sauropodomorph used in this study

recent phylogenetic analyses include numerous "prosauropods" and sauropods, we use geometric morphometrics and multivariate statistics to examine the macroevolutionary history of craniofacial form within this clade, assessing whether broad statements concerning craniofacial diversity are defensible.

Materials and Methods

The sauropodomorph cranial specimens analysed herein are all based upon photographs or reconstructions from the literature (see Larvan 2007 for the specimen list). The landmarks (Fig. 11.2, Table 11.1) were digitised using Image J vers. 1.36b (Rasband 2007).

The program tpsRelw vers. 1.42 (Rohlf 2005) was used to perform the relative warps analyses (RWA; Rohlf 1993). The first step was to perform a generalised Procrustes analysis on the landmark configurations taken from the digital images, to remove differences in location, orientation and scale (see Stayton and Ruta 2006). Once the landmarks were in Procrustes superimposed space a consensus plot (= the mean shape) was calculated. This consensus plot is the mean landmark configuration from all species in the sample. A principle components analysis was done on shape co-ordinates (see Cavalcanti 2004 for the full methodology in detail).

Going from Procrustes superimposed alignment to an ordination plot involves projecting the data points (the taxa) from a curved surface on to a flat one. This can be thought of as taking one side of a globe and projecting the towns and cities onto a map. The process of projection can lead to significant distortion of the relative positions between taxa. To test for this, the Procrustes distances (distances between specimens in Procrustes alignment) were regressed against the Euclidean distances (distances between specimens on the morphospace plot) using the program tpsSmall vers. 1.20 (Rohlf 2003).
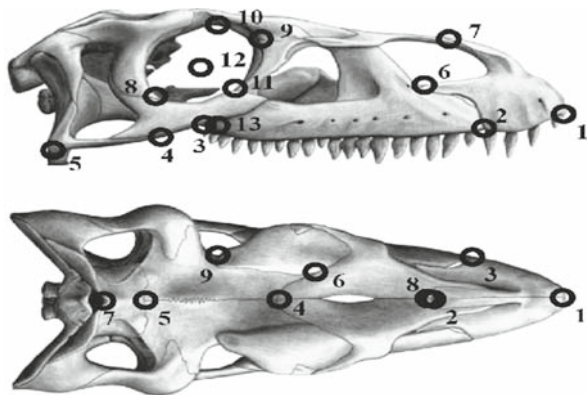
Fig. 11.2 Lateral (*top*) and dorsal (*bottom*) view of a *Melanorosaurus* skull with the landmarks taken shown (see Table 11.1) (image modified from Yates (2007)

Table 11.1 Morphometric landmarks used, with their description

Number	Skull lateral aspect	Skull dorsal aspect
1	Anterior most point along premaxilla ventral margin	Anterior most point of premaxilla along the midline
2	Premaxilla-maxilla suture along ventral margin	Posterior most point of premaxilla along the midline
3	Posterior most point along maxilla ventral margin	Premaxilla-maxilla suture along lateral margin
4	Anterior most point along quadratojugal ventral margin	Nasal-frontal suture along the midline
5	Inflexion point of quadratojugal at the mandibular joint	Frontal-parietal suture along the midline
6	Premaxilla-maxilla suture along dorsal margin	Prefrontal-nasal-frontal suture
7	Dorsal most point of the premaxilla	Posterior most point of the parietal along the midline
8	Jugal-postorbital suture along the orbit rim	Anterior most point of the nasal along the midline
9	Lacrimal-prefrontal suture along the orbit rim	Prefrontal-frontal suture along the orbital margin
10	Prefrontal-frontal suture along the orbit rim	
11	Jugal-lacrimal suture along orbit rim	
12	Centroid of orbit	
13	Posterior end of tooth row	

Four disparity metrics were calculated using all the RW scores from all axes, for "prosauropods" and sauropods: the sum and product of the ranges and variances (Wills et al. 1994). Each metric gives an indication of volume of morphospace occupied. However, range measures quantify the entire spread of morphological variation, or the "absolute extent of bodyplan variety" (Wills 1998), whereas variance measures indicate average dissimilarity among forms. The former are more sensitive to sample size, whereas the latter are more sensitive to taxonomic practice but robustly insensitive to sample size (Wills et al. 1994). All metrics were calculated using the software program Rare (Wills 1998), and multiplicative measures were normalised by taking the 13th root. Statistical significance between the disparity of "prosauropods" and sauropods were assessed in two ways: by the overlap or non-overlap of 95% bootstrap confidence intervals for each disparity metric (calculated by Rare with 1,000 replications) and NPMANOVA (non-parametric multivariate analysis of variance), which tests for significant differences in the distribution of groups in morphospace (Anderson 2001). NPMANOVA, the multivariate (and non-parametric) equivalent of ANOVA, was calculated in PAST (vers. 1.78; Hammer et al. 2001). One of the strengths of NPMANOVA is that as a non-parametric statistical test, it does not assume or require normality from the multivariate data.

In order to assess the fit between RW skull morphspace and phylogeny, Mantel's test was used. From the phylogeny a simple symmetrical node-based difference matrix was collated, while Procrustes distances were calculated using tpsSmall (v1.20, Rohlf 2003) and then collated to a similar symmetrical matrix. The resulting matrices were imported into PAST (v1.78) and a Mantel's test carried out.

Analyses

Regression of Procrustes distance against Euclidean distance for each pair of landmarks gave a high correlation coefficient with $r > 0.99$ for both cranium lateral and cranium dorsal aspect views. As such, distortion did not prevent interpretation of results from the Relative Warp morphospaces.

The first two RW axes in lateral view were deemed significant as they account for 97.7 percent cumulatively of the variance about the mean shape (Fig. 11.3). Moving across RW1, negative to positive, all landmarks move rostrally; 5 (inflexion point of quadratojugal at the mandibular joint), 6 (premaxilla-maxilla suture along dorsal margin) and 7 (dorsal-most point of the premaxilla) only slightly; whereas 8 (jugal-postorbital suture along the orbit rim) and 10 (prefrontal-frontal suture along the orbit rim) "move" considerably. In addition, landmarks 1 (anterior-most point along
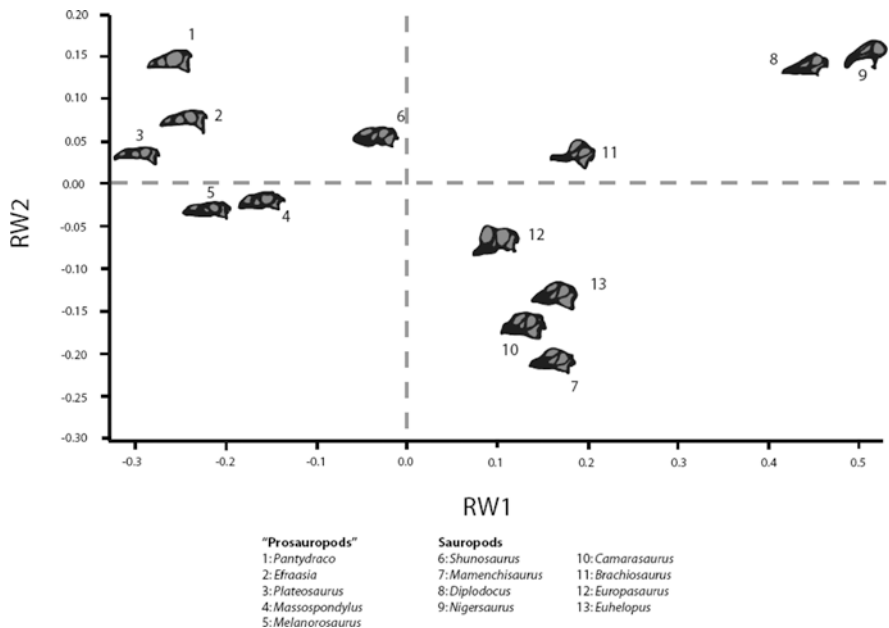
Fig. 11.3 Relative warps lateral view cranial morphology morphospace delimited by the first two axes

the premaxilla ventral margin), 2 (premaxilla-maxilla suture along ventral margin) and 12 (centroid of orbit) exhibit a downwards movement, while landmark 2 converges on landmark 1. Landmark 4 (anterior-most point along quadratojugal ventral margin) is the only one to show an elevation in position. Based on the landmark position changes, progressively positive RW1 values have a lengthened premaxilla and the quadratojugal elongating and advancing rostrally (rostral movement of the jaw joint). The second RW axis subsumes three shape variations: 1) snout elongation, 2) tooth-row shortening and 3) orbit moving up and backwards. Moving negative to positive across this axis, landmarks 1–3 (posterior-most point along maxilla ventral margin) and 8–12 are elevated, 4–7 and 13 (posterior end of tooth row) are depressed, while landmarks 1, 4, 5 and 8–12 "move" caudally and 2, 3, 6, 7 and 13 "move" rostrally.

The morphospace delimited by the first two RW axes in dorsal view (Fig. 11.3) displays separation between "prosauropods" and sauropods across RW1, with *Shunosaurus* intermediate between the two. The diplodocoids (*Diplodocus* and *Nigersaurus*) are distinguished from all other sauropodomorphs, with an elongate snout and a retracted external nares (bony hole for the soft tissue of the nostrils and olfactory senses).

The first two RW axes in dorsal view were deemed significant as they accounted for 72.6 percent cumulatively of the variation about the mean shape (Fig. 11.4). Moving across RW axis one, from negative to positive, landmark 3 (premaxilla-maxilla suture along lateral margin) advances rostrally. Landmarks 2 (posterior-

"Prosauropods"	Sauropods	
1: *Pantydraco*	6: *Shunosaurus*	10: *Camarasaurus*
2: *Efraasia*	7: *Mamenchisaurus*	11: *Brachiosaurus*
3: *Plateosaurus*	8: *Diplodocus*	12: *Europasaurus*
4: *Massospondylus*	9: *Nigersaurus*	13: *Euhelopus*
5: *Melanorosaurus*		
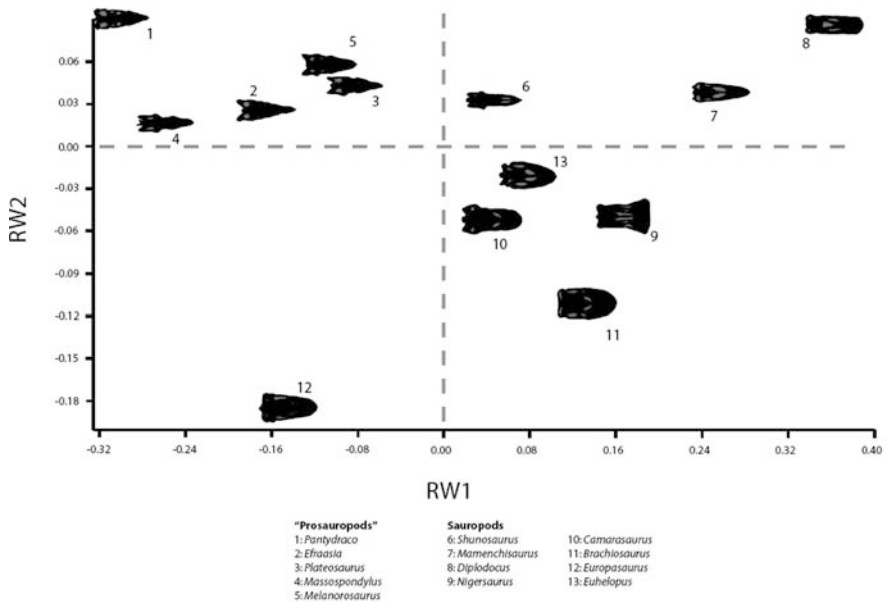
Fig. 11.4 Relative warps dorsal view cranial morphology morphospace delimited by the first two axes

most point of premaxilla along the midline), 4 (nasal-frontal suture along the midline) 6 (prefrontal-nasal-frontal suture), 8 (anterior-most point of the nasal along the midline) and 9 (prefrontal-frontal suture) all "move" caudally. The second RW axis subsumes skull narrowing and lengthening, in addition to narrowing of the premaxillae and the shortening of the nasals. Moving from negative to positive across the axis, landmarks 1 (anterior-most point of the premaxilla along the midline), 4, 5 (frontal-parietal suture along the midline), 6 and 9 are in a rostral position; whereas landmarks 2 and 8 are in a caudal position, and landmarks 3 and 9 "move" in to a medial position. Landmark 7 (posterior-most point of the parietal along the midline) does not alter position in either RW axes.

The morphospace delimited by the first two RW axes in dorsal view (Fig. 11.4) displays separation between "prosauropods" and sauropods across RW1 and RW2, with *Shunosaurus* intermediate between the two. Unlike the lateral view morphospace, the diplodocoids (*Diplodocus* and *Nigersaurus*) are not united as *Nigersaurus* has a laterally expanded snout. *Europasaurus* (the only known case of insular dwarfism within Sauropoda) is the only sauropod with negative RW1 values.

Comparing the distributions of "prosauropods" and sauropods morphospace occupation in both lateral and dorsal view achieved statistical significance (Table 11.2), meaning they occupied significantly distinct regions of morphospace.

Table 11.2 Results from NPMANOVA comparisons of group morphospace distributions from the RW scores. All results based upon 50,000 permutations. The significant p-values are in bold

"View"	Comparison	F-value	P-value
Cranium dorsal	"Prosauropod" Sauropod	4199	**0.0013**
Cranium lateral	"Prosauropod" Sauropod	11.93	**0.0006**

All disparity metrics show that sauropods are more disparate in terms of volume of morphospace occupation; however the very small sample size ("prosauropod" $n = 5$, sauropod $n = 8$) of this analysis severally limits the rarefaction profiles (Fig. 11.5). As the 95% confidence intervals consistently overlapped, statistically significance could not be achieved.

The fit between craniofacial form and phylogeny was found to be statistically significant for both dorsal and lateral view (Table 11.3). However, the correlations themselves were not significant for either view (the highest R^2 being 0.55).

Discussion

This study was primarily concerned with using a measure of morphological variation in order to quantify craniofacial macroevolutionary trends. Relative Warp Analysis was used to measure the variation in skull geometry as the deviation from a
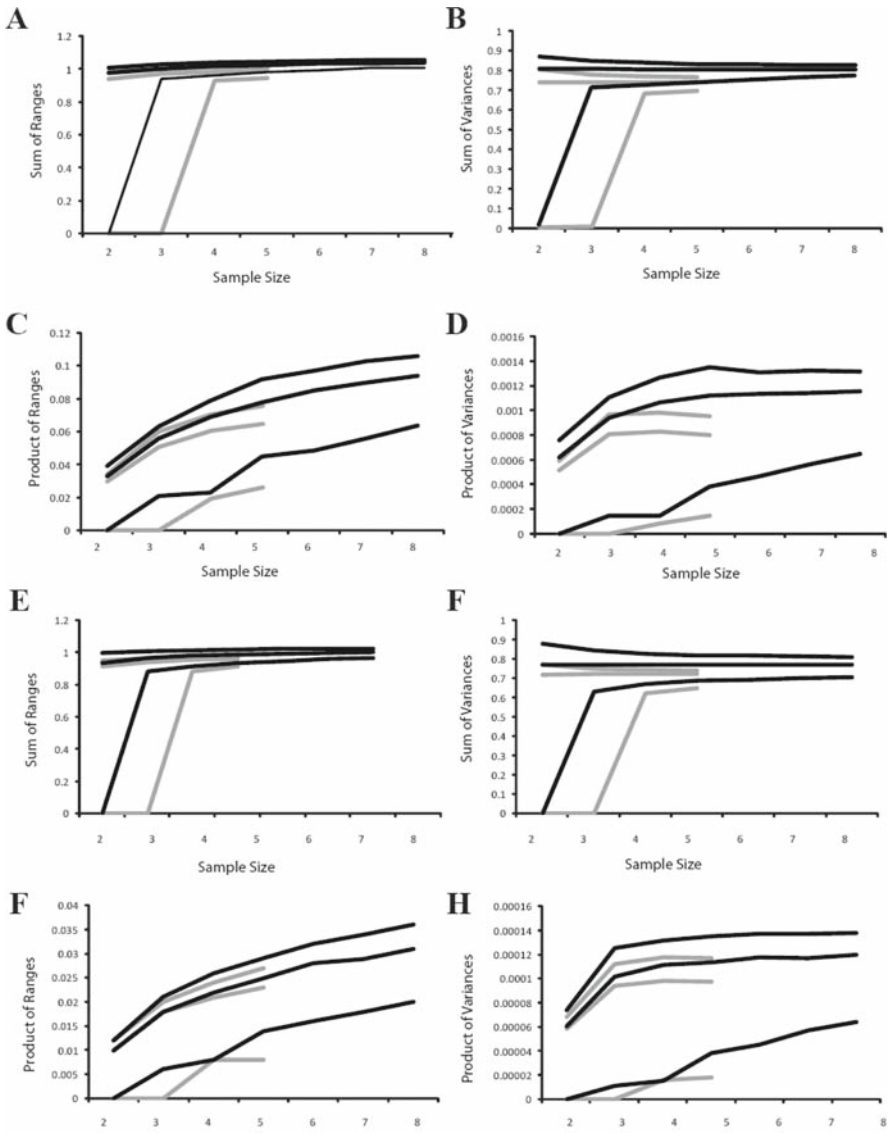
Fig. 11.5 Rarefaction profiles for all four disparity metrics (sum and product of ranges and variances) for "prosauropods" (in *grey*) vs sauropods (*black*). The profiles for the skull dorsal view (**a–d**) and lateral view (**e–h**) are shown with the mean value and 95% confidence intervals as three separate lines

mean shape of the landmark positions. Herein, geometric morphometric techniques delineated "prosauropods" and sauropods in morphospace, although small sample size meant that quantifying volume of morphospace occupation (disparity metrics) failed.

Table 11.3 Correlation results from the Mantel's test of the Procrustes and phylogenetic distances. All results based upon 5,000 permutations. The significant p-values are in bold

"View"	R^2	P-value
Cranium dorsal	0.35	**0.01**
Cranium lateral	0.55	**<0.01**

Relative warps morphospaces for both lateral and dorsal view of the cranium could distinguish clades within Sauropodomorpha. The separation between "prosauropods" and sauropods can be distinguished in their cranial morphology. The prosauropods have lower, more tapered snouts, narrowing towards the tip, whereas sauropods have taller skulls with more robust snouts. The dorsal view shows this varied design, with a clear distinction of "prosauropods", but a large degree of variation within Sauropoda. The basal sauropods (*Shunosaurus* and *Mamenchisaurus*) have a similarly slender appearance, whereas the neosauropods (all taxa more derived than those two, see Fig. 11.1) exhibit a broadening in the snout, resulting in a generally even width from posterior to anterior end of the skull.

The comparatively small morphospace occupied by the "prosauropods" provides very tentative support for the hypothesis that they were more conservative in craniofacial form (and presumably function) and that the sauropods were much more diverse, as shown by their greater morphospace occupation.

With the exception of the dorsal view, diplodocoids are significantly separated in morphospace from all other sauropodomorphs. This is due to their distinct craniofacial form (e.g. elongate snout, retracted external nares), which is possibly related to the feeding strategies hypothesised for this clade (branch stripping and precision biting; Barrett and Upchurch 1994, 2007; Calvo 1994).

Although broad trends within sauropodomorph craniofacial evolution can be discerned, the low correlation between form and phylogeny suggests that phylogenetic inertia has a limited impact upon craniofacial form. This refers to the notion that constraints imposed by phylogenetic history could potentially refrain the skull from reaching its functional optimal.

Conclusions

Geometric morphometrics analyses of skull landmark configurations (lateral and dorsal view) were able to delineate sauropods from "prosauropods". Both views demonstrate that sauropodomorphs became more divergent in terms of craniofacial form throughout the Mesozoic, most especially the sauropods. A clear evolutionary trend towards cranial robustness in sauropods is supported by morphometrics, which also supports the general craniofacial conservativeness of the "prosauropods". Evaluating statistical significance of the distribution and disparity for sauropods and "prosauropods" demonstrate that sauropods occupy a distinct

region of morphospace, and have greater dispersal within morphospace. In addition, there is a statistically significant correlation between craniofacial form and phylogeny; however the correlations themselves are not highly significant.

Although this macroevolutionary study has begun the quantitative elucidation of sauropodomorph evolution, the key problem to our continuing understanding of sauropodomorph craniofacial evolution is the scarcity of well-preserved skulls. Recent studies have shown not only the unrealised cranial diversity of sauropodomorphs (e.g. Sereno et al. 2007), but tooth-crown gross morphology is also highly diverse (e.g. Upchurch and Barrett 2000) and that osteological correlates for structures like a fleshy cheek are known for "prosauropods" and early sauropods (Barrett and Upchurch 2007). Clearly, there are many more exciting discoveries to be made.

References

Anderson MJ. (2001) A new method for non-parametric multivariate analysis of variance. Austral Ecology 26: 32–46.

Barrett PM, Upchurch P. (1994) Feeding mechanisms of *Diplodocus*. Gaia 10: 195–203.

Barrett PM, Upchurch P. (2007) The evolution of feeding mechanisms in early sauropodomorph dinosaurs. In: Barrett PM, Batten DJ (Eds.) Evolution and paleobiology of early sauropodomorph dinosaurs. Special Papers in Palaeontology 77: 91–112.

Calvo JO. (1994) Jaw mechanics in sauropod dinosaurs. Gaia 10: 183–193.

Cavalcanti MJ. (2004) Geometric morphometric analysis of head shape variation in four species of hammerhead sharks (Carcharhiniformes: Sphyrnidae). In: Elewa AMT. (Ed.) Morphometrics: applications in biology and paleontology. Springer-Verlag, Heidelberg: 97–114.

Hammer Ø, Harper DAT, Ryan PD. (2001) PAST: Palaeontological Statistics software package for education and data analysis. Palaeontologia Electronica 4(1): 9.

Henderson DM. (1998) Skull and tooth morphology as indicators of niche partitioning in sympatric Morrison Formation theropods. In: Perez-Moreno BP, Holtz T, Sanz JL, Moratalla J, (Eds.) Aspects of theropod palaeobiology. Gaia 15: 219–226.

Larvan MD. (2007) A combined geometric morphometrics and finite element analysis approach to comparative cranial mechanics of sauropodomorph dinosaurs. Unpublished MSc thesis, University of Bristol.

Rasband W. (2007) Image J v. 1.36b, freeware available from http://rsb.info.nih.gov/ij/

Rohlf FJ. (1993) Relative warp analysis and an example of its application to mosquito wings. In: Marcus LF, Valdecasas A, Bello E. (Eds.) Contributions to morphometrics. Madrid Museum of Natural History, Madrid, Spain: 131–159.

Rohlf FJ. (2003) tpsSmall v.1.20, freeware available from http://life.bio.sunysb.edu/morph/

Rohlf FJ. (2005) tpsRelw v. 1.42, freeware available from http://life.bio.sunysb.edu/morph/

Sander PM. (2000) Longbone histology of the Tendaguru Sauropods: implications for growth and biology. Paleobiology 26: 466–488.

Sereno PC, Wilson JA, Witmer LM, Whitlock JA, Maga A, Ide O, Rowe TA. (2007) Structural extremes in a Cretaceous dinosaur. PLoS ONE 11: e1230.

Stayton CT, Ruta M. (2006) Geometric morphometrics of the skull roof of stereospondyls (Amphibia: Temnospondyli). Palaeontology 49(2): 307–337.

Upchurch P, Barrett PM. (2000) The evolution of sauropod feeding mechanisms. In: Sues H-D. (Eds.) Evolution of herbivory in terrestrial vertebrates: perspectives from the fossil record. Cambridge University Press, Cambridge: 79–122.

Upchurch P, Barrett PM. (2005) Phylogenetic and taxic perspectives on sauropod diversity. In: Rogers KC, Wilson JA (Eds.) The sauropods: evolution and paleobiology. University of California Press, Berkeley: 104–124.

Upchurch P, Barrett PM, Xijin Z, Xing X. (2007) A re-evaluation of *Chinshakiangosaurus chunghoensis* Ye *vide* Dong 1992 (Dinosauria, Sauropodomorpha): implications for cranial evolution in basal sauropod dinosaurs. Geological Magazine 144(2): 247–262.

Wills MA. (1998) Crustacean disparity through the Phanerozoic: comparing morphological and stratigraphic data. Biological Journal of the Linnean Society 65: 455–500.

Wills MA, Briggs DEG, Fortey RA. (1994) Disparity as an evolutionary index: a comparison of Cambrian and Recent arthropods. Paleobiology 20: 93–131.

Yates AM. (2003) A new species of the primitive dinosaur *Thecodontosaurus* (Saurischia: Sauropodomorpha) and its implications for the systematics of early dinosaurs. Journal of Systematic Palaeontology 1: 1–42.

Yates AM. (2007) The first complete skull of the Triassic dinosaur *Melanorosaurus* Haughton (Sauropodomorpha: Anchisauria). In: Barrett PM, Batten, DJ. (Eds.) Evolution and paleobiology of early sauropodomorph dinosaurs. Special Papers in Palaeontology 77: 9–55.

Yates AM, Kitching JW. (2003) The earliest known sauropod dinosaur and the first steps towards sauropod locomotion. Proceedings of the Royal Society of London B 270: 1753–1758.

Chapter 12
The Use of Geometric Morphometrics in Studying Butterfly Wings in an Evolutionary Ecological Context

Casper J. Breuker, Melanie Gibbs, Stefan Van Dongen, Thomas Merckx, and Hans Van Dyck

Idea and Aims

In order to quantify shape variation, many powerful, free, easy-to-use and dedicated software packages have been developed that quickly digitize and/or analyse landmark data (e.g. the tps suite by Rohlf, http://life.bio.sunysb.edu/morph/). Furthermore, powerful shape analyses can also be carried out in statistical packages such as R (Claude 2008). Unlike 10–20 years ago, these days it is therefore no longer difficult to accurately record the position of landmarks on any biological structure. Furthermore, such landmark configurations can easily be compared within and between species using a variety of analyses that together comprise the field of geometric morphometrics (Zelditch et al. 2004). The theoretical core of geometric morphometrics has been well described and is easy to understand, and as such geometric morphometrics can easily be implemented in a wide variety of research fields, such as evolutionary ecology (Zelditch et al. 2004). Using the speckled wood butterfly *Pararge aegeria* (L.) (Nymphalidae, Satyrinae) as our model species, we will illustrate the variety of uses to which geometric morphometrics can be applied to understand the effects of the environment on possibly adaptive butterfly wing size and shape variation in ecologically relevant contexts.

Butterfly Wings and Their Function

In Pterygota (i.e. winged insects), fitness is strongly affected by flight performance (Speight et al. 2008). Behavioural activities such as locating oviposition and foraging sites, courtship, and predator avoidance are often highly dependent on an efficient and precise ability to fly. The design of flight morphology in winged insects is therefore expected to be under strong selection for these behavioural

C.J. Breuker (✉)
Evolutionary Developmental Biology Research Group, Sinclair Building, School of Life Sciences, Oxford Brookes University, Gipsy Lane, Headington, Oxford OX3 0BP, UK
e-mail: cbreuker@brookes.ac.uk

activities which involve staying aloft, manoeuvrability and thrust production (Betts and Wootton 1988; Dudley 2000). A higher wing loading (total body weight/total wing area) (cf. Betts and Wootton 1988) and wing aspect ratio (mean forewing length^2/mean forewing area), for example, would indicate the capacity to fly faster and more extensively (Betts and Wootton 1988). This is, of course, a somewhat simplified view as studies on flight performance in butterflies have shown that flight speed and pattern also depend on factors such as thoracic mass (principally flight muscles), wing damage, centre of body mass, behavioral adaptations (e.g. predation avoidance), thermoregulation, wing asymmetry, and the ability to make use of spatial and temporal variation in the prevailing winds (Srygley and Chai 1990; Srygley and Dudley 1993; Srygley et al. 1996; Srygley 1999; Dudley 2000; Srygley 2000; Srygley and Kingsolver 2000; Srygley 2001, 2003; Berwaerts and Van Dyck 2004).

Male and female butterflies have different life-histories and differ in associated behaviour (Wiklund 2003). Males, for example, spend a significant portion of their time in pursuit of obtaining matings, either by patrolling or perching. Females typically focus on the search for suitable foraging and oviposition sites. Therefore, behavioural sexual dimorphism is often associated with differences in flight morphology and design of the flight apparatus (Breuker et al. 2007a).

Butterfly wings are also used for signalling (Breuker et al. 2006a). The development of the wing pattern elements is to some extent integrated in overall wing development (see later in this chapter). This means that butterfly wings experience a multitude of different natural and sexual selection pressures either simultaneously or sequentially, which will select for a particular wing size and shape in response to the environmental and ecological conditions experienced. These conditions are very likely to be heterogeneous across time and space. Often, what is actually not clear is how this spatio-temporal variation in environmental conditions affects the phenotypic variation of flight morphology.

The Concept of Developmental Plasticity in Ecology

Individuals of the same butterfly species may be able to develop and live in a variety of habitats, or are forced to do so through habitat fragmentation (Hill et al. 2001, 2002; Dover and Settele 2009). Furthermore, the conditions that individuals encounter in these different habitats are very likely to change over time. Habitats may change either suddenly (e.g. a natural catastrophic event), on a regular and more or less predictable basis (e.g. seasonal variation), or progressively (e.g. due to global warming) (Vitousek et al. 1997; Hill et al. 2002). In order to track adaptive peaks, (local) genetic adaptation of flight morphology may occur, which tends to be a relatively slow process as this may take quite a number of generations. In perfectly constant environments, selection is likely to favour just one optimal wing size and shape. Given that most environments, however, are heterogeneous in both space and time, it may be more advantageous to allow for flexibility in wing development, or traits in general, and thus to be able to develop the most optimal flight design in each of the different environments (Bradshaw 1965; Stearns and Koella

1986; Stearns 1989; Schlichting and Pigliucci 1998). Selection may thus operate on the plasticity of traits relevant for survival and reproduction such as flight morphology, thereby ensuring that the development of these traits is flexible enough to allow for a quick response to varying conditions across space and time (Schlichting and Pigliucci 1998; Sultan 2003; West-Eberhard 2003; Sultan 2004). Selection for such developmental flexibility in the case of seasonal variation, for example, is highly likely when selection pressures are different between seasons (Shapiro 1976; Tauber et al. 1986).

This characteristic of development to respond flexibly and adaptively to environmental variability has been labeled developmental plasticity (for an overview of different views on the concept of plasticity, especially concerning its adaptedness, see West-Eberhard 2003), while the resulting phenotypes display a so-called phenotypic plasticity (Bradshaw 1965; Schlichting and Pigliucci 1998; Stearns 2000). Developmental plasticity is thus the modification of development by the environment. This may involve changes in the dynamics of the genetic regulatory network underlying development, thereby resulting in developmental switches (i.e. *developmental conversion* (Smith-Gill 1983)). In general, a set of phenotypes produced by a single genotype across a range of environmental conditions is known as a reaction norm (Woltereck 1909; Pigliucci 2005). A reaction norm with a non-zero slope indicates a plastic response (Fig. 12.1). An example would thus be a range of different wing sizes and shapes across a season or seasons, as for example has been found in damselflies (Bots et al. 2009). These reaction norms may be gradual (i.e. continuous plasticity) or stepwise (i.e. discrete plasticity, as in many examples of seasonal plasticity) (Schlichting and Pigliucci 1998). When individual reaction norms cross, there is genetic variability in the effect environmental heterogeneity has on development (i.e. so-called Genotype by Environment (GxE) effects). GxE effects are commonly observed in life-history traits (Stearns 1989; Via et al. 1995) and are thought to indicate that plasticity can evolve through selection (Schlichting 1989; Moran 1992).

Reaction norms commonly cross, and it is therefore often assumed that most, if not all, observed developmental plasticity is adaptive, and that plasticity is a trait that has been selected for (Scheiner and Lyman 1991). This need not be

Fig. 12.1 Theoretical example of inter- and intra-population trait value differences. Reaction norms of five population 'A' families and five population 'B' families. Population A and B differ in both trait mean and variation. Population A displays no plasticity (hence no GxE)
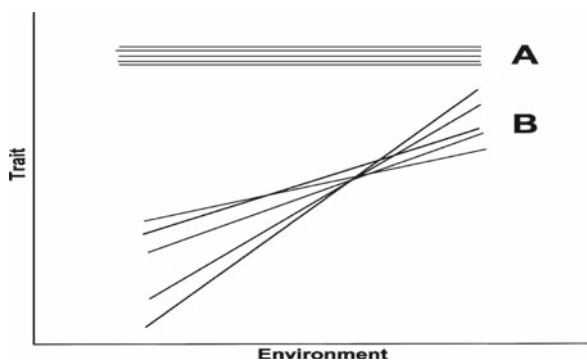
the case, however, as the observed plasticity may simply be a non-adaptive by-product of the physiology and biochemistry of an organism interacting with the variable environment during development (Smith-Gill 1983; Via et al. 1995), with no genetic basis whatsoever (i.e. *developmental modulation* (Smith-Gill 1983)). Developmental conversions are therefore the most likely candidates for adaptive developmental plasticity, as they are presumed to have a genetic basis whereas developmental modulations do not (Smith-Gill 1983). A major complication of GxE studies in a classical quantitative genetic context is the fact that developmental modulation results in, sometimes crossing, reaction norms not unlike those resulting from developmental conversions of the gradual kind. In fact various researchers, for example Smith-Gill (1983), actually lump all *continuous* reaction norms under developmental modulation, and label it non-adaptive. This, however, is considered too restrictive by many (West-Eberhard 2003). There are various ways of testing whether an observed continuous reaction norm is of the developmental conversion or modulation kind, including estimating a selection surface or performing reciprocal transplant experiments. So far, however, the best understood cases of adaptive plasticity remain those where alternative phenotypes fall into discrete classes (e.g. seasonal polyphenism (Shapiro 1976) or wing polyphenism in ants (Abouheif and Wray 2002; Nahmad et al. 2008)).

Developmental and Phenotypic Plasticity in *Pararge aegeria* Wings

Even in the case of seasonal polyphenism, relatively few attempts have been made to explicitly test the hypothesis that the seasonal forms are adapted to the season, and hence the environment individuals live in (Brakefield and Frankino 2009). Butterflies, however, including our model species *P. aegeria*, have been reasonably well-studied for their (adaptive) developmental and phenotypic plasticity of life-history traits and body morphology across space (i.e. different habitats and in relation to habitat fragmentation) and time (mainly seasonal variability in habitat conditions) (Nylin et al. 1989; Van Dyck and Matthysen 1999; Fric and Konvicka 2002; Norberg and Leimar 2002; Van Dyck and Wiklund 2002; Karlsson and Van Dyck 2005; Fric et al. 2006; Breuker et al. 2007a).

Developmental plasticity in *P. aegeria* across landscapes and in relation to habitat fragmentation is still poorly understood. Variation in a number of life-history traits and flight morphology is associated with landscape variability and habitat fragmentation (e.g. Karlsson and Van Dyck 2005), but whether this variation necessarily indicates adaptive plastic responses or merely local adaptation as a result of geographical isolation is unknown as formal tests assessing fitness curves for the plastic traits across landscapes are largely lacking. Merckx and Van Dyck (2006) recently addressed this issue by reciprocally transplanting larvae across different landscapes in a split-brood design. They found that body morphology and flight design showed landscape specific plasticity rather than local adaptation. However, one aspect of

flight design did not appear to be plastic: wing shape, measured as forewing aspect ratio. Nevertheless, this trait does show significant seasonal plasticity (Van Dyck and Wiklund 2002). In another respect this is a somewhat puzzling result as wing shape tends to be very sensitive to environmental variation in insects (Azevedo et al. 1998; Klingenberg et al. 2001; Hoffmann et al. 2002).

It should be noted that none of these *P. aegeria* evolutionary ecological studies used landmarks for the wing shape analyses, which we feel is somewhat of a shortcoming as geometric morphometrics allows one to quantify more subtle wing shape variation. To illustrate how geometric morphometrics can be applied in wing plasticity studies in an evolutionary ecological context, we present an outline of the methods to be used and the results of a recent and unpublished preliminary study.

In this study we tested the hypothesis put forward by Merckx and Van Dyck (2006) that variation in forewing shape in relation to landscape variability is predominantly environmentally determined (i.e. due to developmental plasticity). It has, however, been hypothesised for butterflies that natural selection has favoured changes in flight morphology due to the altered costs and benefits of dispersal in more fragmented landscapes (Hill et al. 1999a, b). Incidentally, studies investigating this hypothesis, for example in *P. aegeria* (Hill et al. 1999a), measured flight morphological traits such as thorax size and relative wing size, but not wing shape (e.g. measured by using a landmark-based approach). We can thus put forward the following hypotheses regarding *P. aegeria* wing shape. If landscapes strongly select for particularly shaped wings, then unrelated populations inhabiting similar landscapes should have similar wing shapes, different from populations from other landscape types, when reared under similar conditions. If, however, wing shape variation is largely due to environmental conditions experienced during development in a particular landscape, then similar rearing conditions are likely to result in a similar wing shape, irrespective of landscape of origin. Rather than performing (large-scale) quantitative genetic experiments in the laboratory to establish the exact role of genetic and environmental variation, and their interaction, on wing morphology (Loh and Bitner-Mathe 2005), we took a somewhat simplified approach in this preliminary study as the experimental design detailed below will make clear.

Quantifying Wing Shape Variation in *Pararge aegeria*

Experimental Design

Speckled wood butterflies were collected from eight different, unrelated, populations in Belgium. Only males were collected. Two representative populations were used from each of four landscape types: (1) large forested woodland (Meerdaalwoud and Bos Ter Rijst), (2) isolated wood (Averechten and Walsbergen), (3) fragmented woodlots (Boshoek and Glabeek), and (4) agricultural landscape (Hoegaarden and Rillaar). The agricultural landscape is a highly fragmented landscape consisting mainly of intensively used fields and pastures, and to a lesser extent of houses,

Table 12.1 The eight different Belgian populations (abbreviations between brackets) – two different populations in each of four different types of landscape – where male speckled wood (*P. aegeria*) butterflies were collected. The number of males (N) is indicated as well as the suitability of the landscape type for *P. aegeria* (1 = poor, 4 = very suitable; ranking is based on Merckx and Van Dyck 2006)

Landscape type	Site	Suitability	N
Large forested woodland	Meerdaalwoud (MDW)	4	17
Large forested woodland	Bos Ter Rijst (BTR)	4	22
Isolated wood	Averechten (AVE)	3	19
Isolated wood	Walsbergen (WAL)	3	18
Fragmented woodlots	Boshoek (BOS)	2	19
Fragmented woodlots	Glabeek (GLA)	2	18
Agricultural landscape	Hoegaarden (HGD)	1	24
Agricultural landscape	Rillaar (RIL)	1	19

farms, orchards and sunken roads with hedgerows. This landscape is largely unsuitable for *P. aegeria* containing no optimal habitat and only an estimated 3% (plus the length of the sunken roads) of suboptimal habitat (Merckx and Van Dyck 2006). Temperatures are highest in this type of landscape compared with the other landscapes in summer (Merckx and Van Dyck 2006; Merckx et al. 2008). For each population 20 individuals were collected, except for Bos Ter Rijst and Hoegaarden, for which 25 individuals were collected. Individuals with excessive wing wear, however, were excluded from the analyses, resulting in the sample size numbers given in Table 12.1.

The males were collected in early spring. This was done deliberately. *Pararge aegeria* butterflies flying in early spring spend the winter in a pupal diapause. Such individuals are very likely to have experienced similar microclimatic environmental conditions during the period that wing development, and thus the position of the wing veins is determined (see below), takes place in the pupa. The direct developing summer generation animals are for example much more likely to experience a range of developmental conditions in the different landscapes. We can thus investigate whether individuals originating from different populations inhabiting different landscapes differ in wing shape under similar rearing conditions in the field. We measured landmarks along the whole of the forewing, at locations where a wing vein meets the edge of the wing (Fig. 12.2), and analysed the data with geometric morphometrics (i.e. Procrustes analyses). For the sake of completion we also investigated forewing length, as this was found to be seasonally plastic in previous studies on *P. aegeria* wing plasticity (Van Dyck and Wiklund 2002).

The forewings were carefully removed from the thorax, placed in between two glass slides and digital images were then taken of the ventral and dorsal wing surface with an Olympus Camedia C-3030 under carefully controlled light conditions (cf. Breuker et al. 2007b). Photographs were randomized with respect to population and landscape type.

Fig. 12.2 Ventral wing surface of a speckled wood *P. aegeria* butterfly. Numbers on the forewing indicate the landmarks measured (i.e. where a wing vein meets the edge of the wing). Forewing length (in mm) is the distance between landmark 1 and 9
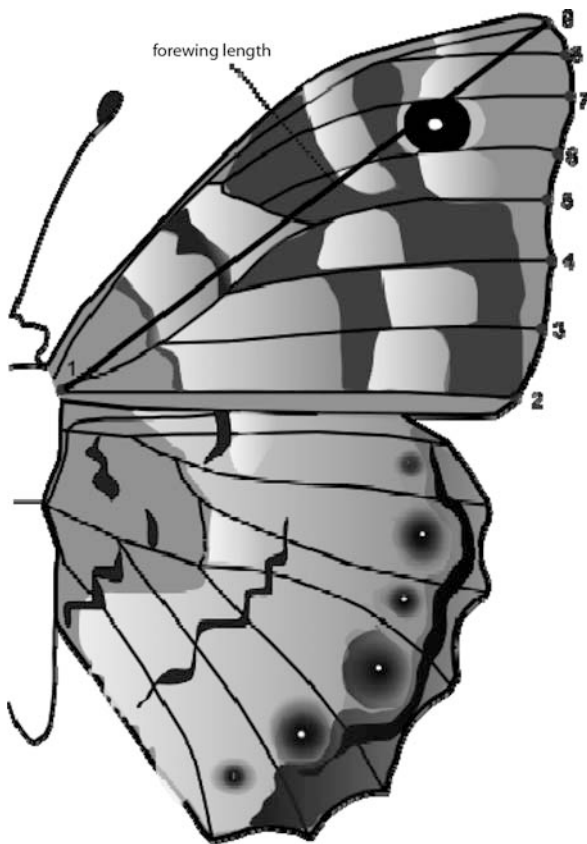
Choosing Landmarks

Given that landmark positions can convey important phylogenetic, developmental and functional information, one should carefully consider which landmarks to digitize and analyze before commencing any morphometrics study (Zelditch et al. 2004). Zelditch et al. (2004) list five criteria to choose landmarks. Landmarks should be (1) homologous, (2) to a degree independent from other landmarks in their location, (3) easy to identify and reliable to measure, (4) lie within the same plane, and (5) provide an adequate coverage of the morphological structure whose shape is to be measured.

In butterflies, and insects in general, the landmarks most often measured are the locations where a wing vein meets the edge of the wing and the wing vein intersections as these landmarks satisfy all five criteria (Dworkin and Gibson 2006; Breuker et al. 2007a, b). Wing vein based landmarks are very easy to identify and measure, and lie in the same plane, as a butterfly wing is a flat structure. In order to fully appreciate how these landmarks satisfy the homology and independence criteria,

and that these landmarks are actually biologically highly relevant when studying wing shape variation, we will briefly explain how insect wings develop.

There are many genes involved in regulating wing size and shape in insects (Mezey and Houle 2005; Mezey et al. 2005), and although many still remain unidentified (De Celis 2003), it has become clear that the basic gene regulatory network underlying wing development is shared among all Pterygota (i.e. winged insects) (Carroll et al. 1994; De Celis 1998; Abouheif and Wray 2002; Carroll et al. 2005). Three key aspects of wing development are integrated: wing vein positioning, cell growth and identity, and intervein cell differentiation (De Celis 1998; Carroll et al. 2005). Decapentaplegic (dpp), a signaling molecule of the TGF-β superfamily, plays a key role in this developmental integration as it regulates growth, patterning and differentiation of wing compartments (Nussbaumer et al. 2000; De Celis 2003; Crozatier et al. 2004; Martin et al. 2004; Dworkin and Gibson 2006; Schwank et al. 2008; Perez-Garijo et al. 2009). Dpp is thus involved in regulating wing size and shape, and the wing venation pattern. Furthermore, the exact identity and position of each individual wing vein is determined progressively in a developing wing disc by an unique combination of up- and down-regulated genes (Nussbaumer et al. 2000; Cook et al. 2004; Hurlbut et al. 2009). In fact, each individual wing vein and intervein region can be regulated independently from other veins and intervein areas (Birdsall et al. 2000; Zimmerman et al. 2000). This independence played a key role in the evolution of species-specific wing venation patterns and wing size and shape (Nussbaumer et al. 2000; De Celis and Diaz-Benjumea 2003). These aspects of wing development satisfy not only the independence criterion, but also the homology criterion. This allows for homologous sets of wing vein based landmarks to be compared within and across species (Breuker et al. 2007b). An example of among species comparisons would be to investigate how the evolution of wing development has created the observed diversity of wing size, shape and wing vein patterning in butterflies.

This means there is one criterion left, that of choosing a set of landmarks to provide adequate coverage of wing shape. In the case of butterflies there are quite a number of landmarks that can be chosen (see the wing venation pattern in Fig. 12.2). Experimental studies that seek to manipulate the transcription and/or translation of genes involved in wing development would need to measure both where a wing vein meets the edge of the wing and wing vein intersections to accurately quantify the effects on the wing venation pattern (Birdsall et al. 2000; Zimmerman et al. 2000; Dworkin and Gibson 2006). In a study such as the one described in this chapter it is sufficient to only measure the locations where a wing vein meets the edge of the wing, and as such capture the ecologically relevant overall "outline" shape of the forewing. We have added landmark 1 (a landmark based on crossing wing veins) to our set of landmarks (Fig. 12.2), as this allowed us to have a better coverage of the overall shape of the wing and have a more reliable measure of wing size (see next section) than would have been obtained with landmarks 2–9 only.

There is one more aspect in choosing landmarks that should be considered here. The fifth criterion speaks of an *adequate* number of landmarks (Zelditch et al. 2004). The lower limit of the number of landmarks is determined by the fact that too few

landmarks would of course not convey enough shape information. The upper limit may to a large extent be determined or constrained by the sample size. Shape is a multivariate trait and in order to have enough degrees of freedom to carry out geometric morphometrics analyses the number of individuals measured should be in excess of twice the number of landmarks minus 4 (Zelditch et al. 2004). Given our sample sizes (Table 12.1), 9 landmarks is definitely the upper limit. Ideally, one should really have larger sample sizes, but for most butterfly species it is difficult to actually catch large numbers of undamaged individuals in a limited period (such as early spring, which is about two weeks) in different sites at the same time. Furthermore, it is not always easy to catch similar numbers of females and males. In this study, for example, insufficient numbers of females were found, as their flight behaviour is much more cryptic than that of males.

Geometric Morphometrics

The 9 landmarks were digitized on the ventral side of the forewing in ImageJ (freely available on http://rsb.info.nih.gov/ij/) (Fig. 12.2). Variation in shape was examined by using geometric morphometrics based on generalized least squares Procrustes superimposition methods (Goodall 1991; Dryden and Mardia 1998; Klingenberg and McIntyre 1998). Procrustes methods analyze shape by superimposing configurations of landmarks of two or more individuals to achieve an overall best fit. It involves four steps, which have been described in mathematical and descriptive detail elsewhere (see details in Klingenberg and McIntyre 1998): (1) reflection of either left or right configurations (i.e. so left and right are now orientated the same way), (2) scaling to unit centroid size (to remove size and shape associations), (3) superimposing the centroids of all configurations, and finally (4) rotation of the configurations around their centroid to obtain the optimal alignment. Step 1 only needs to be performed when wings from both sides are used.

To estimate the amount of measurement error, and thereby to ensure that measurement error due to imaging and digitizing was negligible compared to biological shape and size variation, repeated photos and measurements were taken for a subset of individuals, and a Procrustes ANOVA (Klingenberg and McIntyre 1998) was performed. This will establish how reliable our 9 landmarks were to measure. Measurement error was found to be negligible as the mean squares for individual, side and asymmetry between the sides (the side × individual interaction) significantly exceeded the mean squares of the error term ($P \ll 0.001$; Table 12.2).

We used the Procrustes Distance to summarize shape differences between the average shape of two populations (Klingenberg and McIntyre 1998). The square root of the sum of the squared distances between corresponding landmarks of two optimally aligned configurations is an approximation of this Procrustes Distance. Differences in shape between populations were analysed by means of canonical variates analysis (CVA) applying 10,000 permutations, which is more than a reasonable number of permutations for the vast majority of shape comparisons. A permutation test is used here to determine statistical significance as a

Table 12.2 Analysis of measurement error in digitizing landmarks on a *P. aegeria* forewing. Procrustes analysis of shape variance (Klingenberg and McIntyre 1998) of the amounts of shape variation attributable to different sources, for a random subset of 103 individuals, which were digitized twice. Nine landmarks were measured. The measurement error consists of both the imaging and digitizing error. Sums of squares and mean squares are in units of squared Procrustes Distance. *** $P < 0.001$

Source	Sums of squares	Degrees of freedom	Mean squares $\times 10^6$
Individuals	0.3211	1428	224.86***
Sides	0.002142	14	152.97***
Sides × Individuals	0.03370	1428	23.60***
Error	0.01966	2884	6.82

permutation test does not need to assume any statistical distribution for shape variation. CVA finds the axes (i.e. canonical variates, CVs) that optimize the between population differences relative to the within population variation in landmark configurations. Patterns of shape differentiation that CVs represent can be plotted graphically. Multiplying the original shape variables by the coefficients of the CVs, and summing them, produces a series of vectors of relative landmark displacement that illustrates the shape differentiation represented by the CVs (Zelditch et al. 2004).

The centroid size of all nine landmarks of the forewing was used in this study as a measure of forewing size. Centroid size is the square root of the sum of squared distances from a set of landmarks to their centroid (i.e. mean x and y coordinate of a set of landmarks per individual) (see e.g. Klingenberg and McIntyre 1998). Forewing length is the distance (in mm) between landmark 1 at the base of the wing, and landmark 9 at the wing tip (Fig. 12.2).

The Procrustes ANOVA to determine the measurement error was performed in SAGE, written by E. Marquez (http://www-personal.umich.edu/~emarquez/morph/). Wing size and forewing length differences between populations were analysed by means of ANOVA in R 2.8.1 (http://cran.r-project.org). Procrustes superimposition and the canonical variates analyses to determine the forewing shape differences between populations and different landscapes were performed in MorphoJ (version 1.00e) (Klingenberg 2008).

Results

Forewing Shape Differences Between Populations and Different Landscapes

The CVA results in Table 12.3 show that male forewing shape differences between any two populations are non-significant. This means that the shape variation within a landscape type (i.e. between the two populations of a landscape) is similar to the variation among landscapes. A scatter plot with the two CVs (CV1 and CV2) that optimized the between population forewing shape differences relative to within

Table 12.3 Comparison among the eight populations for male *P. aegeria* forewing shape by means of Canonical Variate Analysis (CVA). Shape variation between two populations is estimated as the Procrustes Distance (Klingenberg and Monteiro 2005). P-values were calculated with 10,000 random permutations per test

	AVE	BOS	BTR	GLA	HGD	MDW	RIL
BOS	0.013 $P = 0.16$						
BTR	0.0076 $P = 0.56$	0.0098 $P = 0.24$					
GLA	0.0089 $P = 0.51$	0.0084 $P = 0.48$	0.0085 $P = 0.39$				
HGD	0.0153 $P = 0.14$	0.0045 $P = 0.94$	0.0119 $P = 0.21$	0.0087 $P = 0.56$			
MDW	0.0089 $P = 0.49$	0.0065 $P = 0.70$	0.0071 $P = 0.53$	0.0056 $P = 0.83$	0.0079 $P = 0.64$		
RIL	0.014 $P = 0.13$	0.005 $P = 0.90$	0.0115 $P = 0.15$	0.008 $P = 0.57$	0.0043 $P = 0.96$	0.0068 $P = 0.69$	
WAL	0.0041 $P = 0.96$	0.010 $P = 0.21$	0.0062 $P = 0.59$	0.0064 $P = 0.64$	0.0121 $P = 0.22$	0.0065 $P = 0.61$	0.0108 $P = 0.19$

population shape variation illustrates this result (Fig. 12.3). Figure 12.4 illustrates the patterns of shape differentiation they represent. None of the CVs, including CV1 and CV2, were effective discriminators between the populations as there are no significant forewing shape differences between the populations.
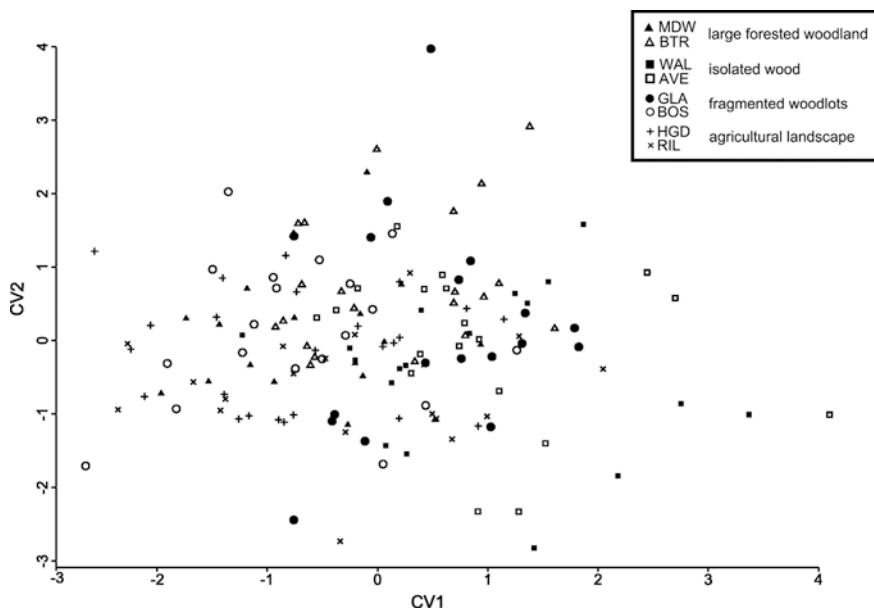
Fig. 12.3 Scatterplot (CV1 versus CV2) from Canonical Variate Analysis of male *P. aegeria* forewing shape from eight different, non-overlapping populations inhabiting 4 different landscapes. Abbreviations of the populations, and more information on them, can be found in Table 12.1

(A)

(B)
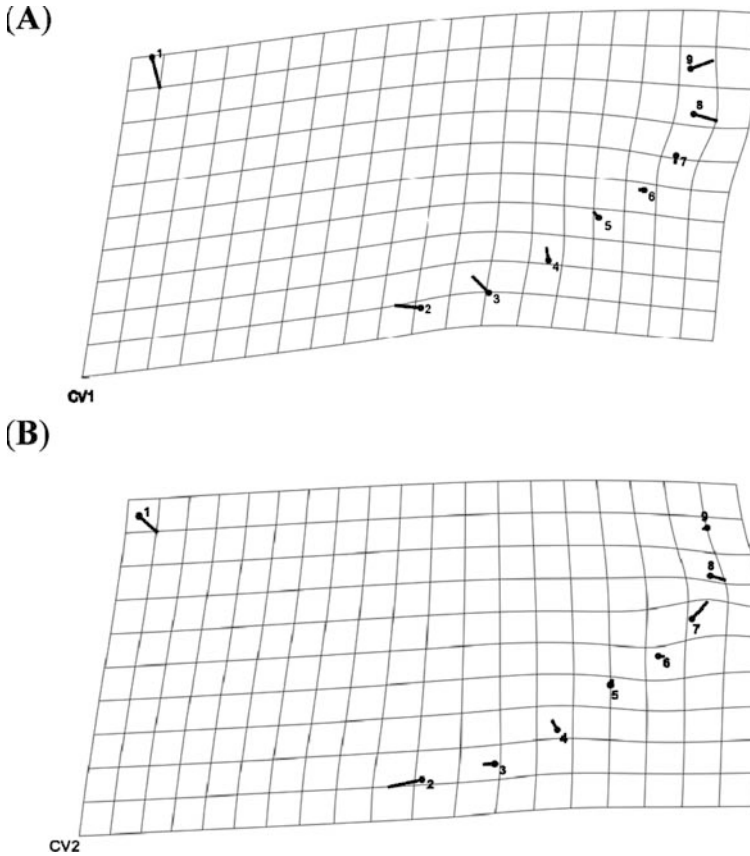
Fig. 12.4 Shape differentiation associated with CV1 (**a**) and CV2 (**b**) for the eight populations from which male *P. aegeria* were collected. The lollipops and deformation grid indicate the directions and magnitude of shape variation along the CV

Wing Size Differences Between Populations

Males from the eight populations did not differ significantly in forewing length (one-way ANOVA: $F_{7,149} = 0.75$, $P = 0.63$), nor did they differ in wing size as estimated by centroid size (one-way ANOVA: $F_{7,149} = 1.28$, $P = 0.26$). Pooling the two populations per landscape type showed exactly the same pattern for forewing length (one-way ANOVA: $F_{3,153} = 0.60$, $P = 0.61$) and centroid size (one-way ANOVA: $F_{3,153} = 1.39$, $P = 0.25$).

Phenotypic Plasticity?

The results seem to indicate that populations of *P. aegeria* butterflies inhabiting different landscapes developed very similar wing shapes and sizes. These results were of course based on small sample sizes, and it is therefore clear that this experiment

needs to be repeated, in both the field and the lab (which is currently being done). It seems, however, that there is an indication that developmental plasticity plays a crucial role in wing morphology in *P. aegeria* butterflies.

Although numerous studies have reported the existence of wing and trait plasticity in general (Van Dyck and Wiklund 2002; Merckx and Van Dyck 2006), investigated the quantitative genetics of plasticity (Winterhalter and Mousseau 2007), and even modeled its evolution (Via and Lande 1985; de Jong 2005), the actual developmental and genetic mechanisms underlying plasticity of fitness-related traits such as wing morphology and its adaptive value in different ecological contexts are still very poorly understood (Marden 2006). Knowing these genetic mechanisms is essential as it will make it possible to determine whether a (continuous) reaction norm is the result of a developmental conversion (and therefore possibly adaptive) or merely developmental modulation. More importantly though, it will unlock many of the secrets underlying the flexibility in gene expression patterns that provide both developmental robustness and plasticity, which are of profound evolutionary significance (Abouheif and Wray 2002; Marden 2006; Nahmad et al. 2008).

As indicated in the section on choosing landmarks, changes in the dynamics of the regulatory network underlying wing development, either by means of careful experimental manipulation of gene expression or induced by the environment, can thus have changes on overall wing size and shape, which can be traced to individual wing veins or compartments (Birdsall et al. 2000; Zimmerman et al. 2000). This can be quantified at the phenotypic level by means of geometric morphometrics, which will indicate which veins changed position (Birdsall et al. 2000; Zimmerman et al. 2000; Breuker et al. 2006b; Dworkin and Gibson 2006). Carefully designed experiments would therefore offer the possibility of elucidating the developmental mechanisms underlying wing shape developmental plasticity observed in a variety of ecological contexts. A fully integrated research programme on adaptive plasticity in butterfly wing morphology would thus need to establish spatio-temporal variation due to environmental regulation of development, and demonstrate that this regulation is adaptive. We argue that geometric morphometrics will also play a crucial role in these developmental studies (cf. Klingenberg and McIntyre 1998; Birdsall et al. 2000; Zimmerman et al. 2000; cf. Dworkin and Gibson 2006).

Summary

In this chapter we have illustrated, using the speckled wood butterfly *P. aegeria* as our model species, the ease with which geometric morphometrics can be applied in evolutionary ecology, using free and easy-to-use morphometric software packages. Great care should be taken in choosing landmarks and multiple criteria should be met when choosing a set of landmarks. Landmarks based on the position of butterfly wing veins do not only meet these criteria, they are also biologically relevant as wing veins play a key role in our understanding of the developmental processes underlying wing shape and its evolution. This makes butterfly wings ideal for the

implementation of geometric morphometrics in a wide variety of studies, not only evolutionary ecological studies, but also developmental studies.

References

Abouheif, E. & Wray, G.A. 2002. Evolution of the gene network underlying wing polyphenism in ants. Science 297: 249–252.

Azevedo, R.B.R., James, A.C., McCabe, J. & Partridge, L. 1998. Latitudinal variation of wing: thorax size ratio and wing-aspect ratio in *Drosophila melanogaster*. Evolution 52: 1353–1362.

Berwaerts, K. & Van Dyck, H. 2004. Take-off performance under optimal and suboptimal thermal conditions in the butterfly *Pararge aegeria*. Oecologia 141: 536–545.

Betts, C.R. & Wootton, R.J. 1988. Wing shape and flight behaviour in butterflies (Lepidoptera: papilionoidea and hesperioidea): a preliminary analysis. Journal of Experimental Biology 138: 271–288.

Birdsall, K., Zimmerman, E., Teeter, K. & Gibson, G. 2000. Genetic variation for the positioning of wing veins in *Drosophila melanogaster*. Evolution & Development 2: 16–24.

Bots, J., Breuker, C.J., Van Kerkhove, A., Van Dongen, S., De Bruyn, L. & Van Gossum, H. 2009. Variation in flight morphology in a female polymorphic damselfly: intraspecific, intrasexual, and seasonal differences. Canadian Journal of Zoology 87: 86–94.

Bradshaw, A.D. 1965. Evolutionary significance of phenotypic plasticity in plants. Advances in Genetics 13: 115–155.

Brakefield, P.M. & Frankino, W.A. 2009. Polyphenisms in Lepidoptera: multidisciplinary approaches to studies of evolution. In: Phenotypic plasticity of insects: mechanisms and consequences (T. N. Ananthakrishnan & D. W. Whitman, eds.), pp. 121–152. Science Publishers, Enfield.

Breuker, C.J., Debat, V. & Klingenberg, C.P. 2006a. Functional evo-devo. Trends in Ecology & Evolution 21: 488–492.

Breuker, C.J., Patterson, J.S. & Klingenberg, C.P. 2006b. A single basis for developmental buffering of *Drosophila* wing shape. PLOS-one 1: e7.

Breuker, C.J., Brakefield, P.M. & Gibbs, M. 2007a. The association between wing morphology and dispersal is sex-specific in the glanville fritillary butterfly *Melitaea cinxia* (Lepidoptera: Nymphalidae). European Journal of Entomology 104: 445–452.

Breuker, C.J., Gibbs, M., Van Dyck, H., Brakefield, P.M., Klingenberg, C.P. & Van Dongen, S. 2007b. Integration of wings and their eyespots in the speckled wood butterfly *Pararge aegeria*. Journal of Experimental Zoology Part B-Molecular and Developmental Evolution 308B: 454–463.

Carroll, S.B., Gates, J., Keys, D.N., Paddock, S.W., Panganiban, G.E.F., Selegue, J.E. & Williams, J.A. 1994. Pattern formation and eyespot determination in butterfly wings. Science 265: 109–114.

Carroll, S.B., Grenier, J.K. & Weatherbee, S.B. 2005. From DNA to diversity: molecular genetics and the evolution of animal design, 2nd edn. Blackwell Publishing, Oxford.

Claude, J. 2008. Morphometrics with R. Springer, New York.

Cook, O., Biehs, B. & Bier, E. 2004. *brinker* and *optomotor-blind* act coordinately to initiate development of the L5 wing vein primordium in *Drosophila*. Development 131: 2113–2124.

Crozatier, M., Glise, B. & Vincent, A. 2004. Patterns in evolution: veins of the *Drosophila* wing. Trends in Genetics 20: 498–505.

De Celis, J.F. 1998. Positioning and differentiation of veins in the *Drosophila* wing. International Journal of Developmental Biology 42: 335–343.

De Celis, J.F. 2003. Pattern formation in the *Drosophila* wing: the development of the veins. Bioessays 25: 443–451.

De Celis, J.F. & Diaz-Benjumea, F.J. 2003. Developmental basis for vein pattern variations in insect wings. International Journal of Developmental Biology 47: 653–663.

de Jong, G. 2005. Evolution of phenotypic plasticity: patterns of plasticity and the emergence of ecotypes. New Phytologist 166: 101–117.

Dover, J. & Settele, J. 2009. The influences of landscape structure on butterfly distribution and movement: a review. Journal of Insect Conservation 13: 3–27.

Dryden, I.L. & Mardia, K.V. 1998. Statistical shape analysis. Wiley, Chichester.

Dudley, R. 2000. The biomechanics of insect flight: form, function and evolution. Princeton University Press, Princeton.

Dworkin, I. & Gibson, G. 2006. Epidermal growth factor receptor and transforming growth factor-beta signaling contributes to variation for wing shape in *Drosophila melanogaster*. Genetics 173: 1417–1431.

Fric, Z. & Konvicka, M. 2002. Generations of the polyphenic butterfly *Araschnia levana* differ in body design. Evolutionary Ecology Research 4: 1017–1032.

Fric, Z., Klimova, M. & Konvicka, M. 2006. Mechanical design indicates differences in mobility among butterfly generations. Evolutionary Ecology Research 8: 1511–1522.

Goodall, C.R. 1991. Procrustes methods in the statistical analysis of shape (with discussion). Journal of the Royal Statistical Society Series B 53: 285–339.

Hill, J.K., Thomas, C.D. & Blakeley, D.S. 1999a. Evolution of flight morphology in a butterfly that has recently expanded its geographic range. Oecologia 121: 165–170.

Hill, J.K., Thomas, C.D. & Lewis, O.T. 1999b. Flight morphology in fragmented populations of a rare British butterfly, *Hesperia comma*. Biological Conservation 87: 277–283.

Hill, J.K., Collingham, Y.C., Thomas, C.D., Blakeley, D.S., Fox, R., Moss, D. & Huntley, B. 2001. Impacts of landscape structure on butterfly range expansion. Ecology Letters 4: 313–321.

Hill, J.K., Thomas, C.D., Fox, R., Telfer, M.G., Willis, S.G., Asher, J. & Huntley, B. 2002. Responses of butterflies to twentieth century climate warming: implications for future ranges. Proceedings of the Royal Society of London Series B-Biological Sciences 269: 2163–2171.

Hoffmann, A.A., Collins, E. & Woods, R. 2002. Wing shape and wing size changes as indicators of environmental stress in *Helicoverpa punctigera* (Lepidoptera: Noctuidae) moths: comparing shifts in means, variances, and asymmetries. Environmental Entomology 31: 965–971.

Hurlbut, G.D., Kankel, M.W. & Artavanis-Tsakonas, S. 2009. Nodal points and complexity of Notch-Ras signal integration. Proceedings of the National Academy of Sciences 106: 2218–2223.

Karlsson, B. & Van Dyck, H. 2005. Does habitat fragmentation affect temperature-related life-history traits? A laboratory test with a woodland butterfly. Proceedings of the Royal Society B-Biological Sciences 272: 1257–1263.

Klingenberg, C.P. 2008. MorphoJ. Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm

Klingenberg, C.P. & McIntyre, G.S. 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with procrustes methods. Evolution 52: 1363–1375.

Klingenberg, C.P., Badyaev, A.V., Sowry, S.M. & Beckwith, N.J. 2001. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings. American Naturalist 157: 11–23.

Klingenberg, C.P. & Monteiro, L.R. 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. Systematic Biology 54: 678–688.

Loh, R. & Bitner-Mathe, B.C. 2005. Variability of wing size and shape in three populations of a recent Brazilian invader, *Zaprionus indianus* (Diptera: Drosophilidae), from different habitats. Genetica 125: 271–281.

Marden, J.H. 2006. Quantitative and evolutionary biology of alternative splicing: how changing the mix of alternative transcripts affects phenotypic plasticity and reaction norms. Heredity, 10.1038/sj.hdy.6800904.

Martin, F.A., Perez-Garijo, A., Moreno, E. & Morata, G. 2004. The *brinker* gradient controls wing growth in *Drosophila*. Development 131: 4921–4930.

Merckx, T. & Van Dyck, H. 2006. Landscape structure and phenotypic plasticity in flight morphology in the butterfly *Pararge aegeria*. Oikos 113: 226–232.

Merckx, T., Van Dongen, S., Matthysen, E. & Van Dyck, H. 2008. Thermal flight morphology budget of a woodland butterfly in woodland versus agricultural landscapes: an experimental assessment. Basic and Applied Ecology 9: 433–442.

Mezey, J.G. & Houle, D. 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution 59: 1027–1038.

Mezey, J.G., Houle, D. & Nuzhdin, S.V. 2005. Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics 169: 2101–2113.

Moran, N.A. 1992. The evolutionary maintenance of alternative phenotypes. American Naturalist 139: 971–989.

Nahmad, M., Glass, L. & Abouheif, E. 2008. The dynamics of developmental system drift in the gene network underlying wing polyphenism in ants: a mathematical model. Evolution & Development 10: 360–374.

Norberg, U. & Leimar, O. 2002. Spatial and temporal variation in flight morphology in the butterfly *Melitaea cinxia* (Lepidoptera: Nymphalidae). Biological Journal of the Linnean Society 77: 445–453.

Nussbaumer, U., Halder, G., Groppe, J., Affolter, M. & Montagne, J. 2000. Expression of the *blistered/DSRF* gene is controlled by different morphogens during *Drosophila* trachea and wing development. Mechanisms of Development 96: 27–36.

Nylin, S., Wickman, P.-O. & Wiklund, C. 1989. Seasonal plasticity in growth and development of the speckled wood butterfly, *Pararge aegeria* (Satyrinae). Biological Journal of the Linnean Society 38: 155–171.

Perez-Garijo, A., Shlevkov, E. & Morata, G. 2009. The role of Dpp and Wg in compensatory proliferation and in the formation of hyperplastic overgrowths caused by apoptotic cells in the *Drosophila* wing disc. Development 136: 1169–1177.

Pigliucci, M. 2005. Evolution of phenotypic plasticity: where are we going now? Trends in Ecology & Evolution 20: 481–486.

Scheiner, S.M. & Lyman, R.F. 1991. The Genetics of phenotypic plasticity. II. Response to selection. Journal of Evolutionary Biology 4: 23–50.

Schlichting, C.D. 1989. Phenotypic integration and environmental change. What are the consequences of differential phenotypic plasticity of traits. Bioscience 39: 460–464.

Schlichting, C.D. & Pigliucci, M. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer Associates, Sunderland.

Schwank, G., Restrepo, S. & Basler, K. 2008. Growth regulation by Dpp: an essential role for Brinker and a non-essential role for graded signaling levels. Development 135: 4003–4013.

Shapiro, A.M. 1976. Seasonal polyphenism. Evolutionary Biology 9: 259–333.

Smith-Gill, S.J. 1983. Developmental plasticity: developmental conversion versus phenotypic modulation. American Zoologist 23: 47–55.

Speight, M.R., Hunter, M.D. & Watt, A.D. 2008. Ecology of insects: concepts and applications, 2nd edn. Wiley-Blackwell, Oxford.

Srygley, R.B. & Chai, P. 1990. Flight morphology of Neotropical butterflies: palatability and distribution of mass to the thorax and abdomen. Oecologia, Berlin 84: 491–499.

Srygley, R.B. & Dudley, R. 1993. Correlations of the position of center of body mass with butterfly escape tactics. Journal of Experimental Biology 174: 155–166.

Srygley, R.B., Oliveira, E.G. & Dudley, R. 1996. Wind drift compensation, flyways, and conserva-
tion of diurnal, migrant Neotropical Lepidoptera. Proceedings of the Royal Society of London
Series B-Biological Sciences 263: 1351–1357.

Srygley, R.B. 1999. Incorporating motion into investigations of mimicry. Evolutionary Ecology
13: 691–708.

Srygley, R.B. 2000. Locomotor mimicry among passion-vine butterflies Heliconius. American
Zoologist 40: 1219–1219.

Srygley, R.B. & Kingsolver, J.G. 2000. Effects of weight loading on flight performance and sur-
vival of palatable Neotropical *Anartia fatima* butterflies. Biological Journal of the Linnean
Society 70: 707–725.

Srygley, R.B. 2001. Compensation for fluctuations in crosswind drift without stationary landmarks
in butterflies migrating over seas. Animal Behaviour 61: 191–203.

Srygley, R.B. 2003. Locomotor mimicry and energetic costs of aposematic signaling in butterflies.
Integrative and Comparative Biology 43: 823.

Stearns, S.C. & Koella, J.C. 1986. The evolution of phenotypic plasticity in life-history traits –
predictions of reaction norms for age and size at maturity. Evolution 40: 893–913.

Stearns, S.C. 1989. The evolutionary significance of phenotypic plasticity – Phenotypic sources of
variation among organisms can be described by developmental switches and reaction norms.
Bioscience 39: 436–445.

Stearns, S.C. 2000. Daniel Bernoulli (1738): evolution and economics under risk. Journal of
Biosciences 25: 221–228.

Sultan, S.E. 2003. Commentary: the promise of ecological developmental biology. Journal of
Experimental Zoology Part B-Molecular and Developmental Evolution 296B: 1–7.

Sultan, S.E. 2004. Promising directions in plant phenotypic plasticity. Perspectives in Plant
Ecology Evolution and Systematics 6: 227–233.

Tauber, M.J., Tauber, C. & Masaki, S. 1986. Seasonal adaptations of insects. Oxford University
Press, Oxford.

Van Dyck, H. & Matthysen, E. 1999. Habitat fragmentation and insect flight: a changing 'design'
in a changing landscape? Trends in Ecology & Evolution 14: 172–174.

Van Dyck, H. & Wiklund, C. 2002. Seasonal butterfly design: morphological plasticity among three
developmental pathways relative to sex, flight and thermoregulation. Journal of Evolutionary
Biology 15: 216–225.

Via, S. & Lande, R. 1985. Genotype-environment interaction and the evolution of phenotypic
plasticity. Evolution 39: 505–522.

Via, S., Gomulkiewicz, R., Dejong, G., Scheiner, S.M., Schlichting, C.D. & Vantienderen, P.H.
1995. Adaptive phenotypic plasticity – consensus and controversy. Trends in Ecology &
Evolution 10: 212–217.

Vitousek, P.M., Mooney, H.A., Lubchenco, J. & Melillo, J.M. 1997. Human domination of Earth's
ecosystems. Science 277: 494–499.

West-Eberhard, M.J. 2003. Developmental plasticity and evolution. Oxford University Press,
New York.

Wiklund, C. 2003. Sexual selection and the evolution of butterfly mating systems. In: Butterflies:
ecology and evolution taking flight (C. L. Boggs, W. B. Watt & P. R. Ehrlich, eds.), pp. 67–90.
Chicago University Press, Chicago.

Winterhalter, W.E. & Mousseau, T.A. 2007. Patterns of phenotypic and genetic variation for the
plasticity of diapause incidence. Evolution 61: 1520–1531.

Woltereck, R. 1909. Weitere experimentelle Untersuchungen über Artveränderung, speziell
über das Wesen quantitativer Artunterschiede bei Daphniden. Verhandlungen der Deutschen
Zoologischen Gesellschaft 19: 110–172.

Zelditch, M.L., Swiderski, D.L., Sheets, H.D. & Fink, W.L. 2004. Geometric morphometrics for
biologists. Elsevier Academic Press, London.

Zimmerman, E., Palsson, A. & Gibson, G. 2000. Quantitative trait loci affecting components of
wing shape in *Drosophila melanogaster*. Genetics 155: 671–683.

Chapter 13
Towards Automating Artifact Analysis: A Study Showing Potential Applications of Computer Vision and Morphometrics to Artifact Typology

Michael J. Lenardi and Daria E. Merwin

Idea and Aims

First released in 1961, William A. Ritchie's *A Typology and Nomenclature of New York State Projectile Points* remains an influential resource for archaeologists working in the northeastern US. Ritchie examined a large and diverse sample of stone projectile points, but not all fit neatly into his typology. In this chapter, we propose that alternative approaches afforded by computer vision and morphometrics can shed light on this and other problems of traditional stone tool typology. With the advent of computer vision, we can now examine the entire morphological continuum of projectile points through statistical shape analysis. Following automated image capture, three analytical methods were evaluated using silhouette, outline, and landmark data. Biased free capture methods to record the form of an individual projectile point, together with using invariant shape descriptors to quantify the data, may result in more objective analysis than was possible in the past.

Introduction

Archaeology is the study of the past that uses material remains to reconstruct human behavior. As archaeologists, we seek to understand the similarities and differences within and between groups, often over large distances in both time and space. Our attempts to understand the past are often framed by broad questions: how and when did human groups spread across the globe? What are the origins, and repercussions, of plant and animal domestication? Why did social complexity develop in some places and not others? The answers to these questions lie in the archaeological record of artifacts and features, the tangible remains of past human activities. However, the data set is far from complete. Many aspects of human behavior simply

M.J. Lenardi (✉)
Cultural Resource Survey Program, New York State Museum, Cultural Education Center, Albany, NY 12230, USA
e-mail: mlenardi@mail.nysed.gov

do not leave significant traces, while other clues to the past typically do not survive in the ground for centuries or millennia, such as fragile organic remains (e.g., objects made from animal or plant materials).

The most ubiquitous class of artifact found on archaeological sites worldwide consists of stone tools and the waste products from their manufacture. Chipped stone implements are found on sites left by some of our earliest hominid ancestors, and are still used today by some groups. Stone artifacts are frequently the only type of artifact recovered from archaeological deposits left by pre-ceramic peoples. Raw material for stone tool manufacture is found in most environments, and it is extremely durable, making it valuable not only for its original user but also for archaeologists.

Archaeologists working in the northeastern United States usually employ a system of three broad periods (Paleoindian, Archaic, Woodland) to divide the span of time between the first settlement of the region by Native peoples and the arrival of the European explorers and colonists in the sixteenth century. Although much fieldwork has taken place, many questions regarding the region's prehistory remain. Interpretation has been hindered by a lack of radiocarbon dates from prehistoric Native American sites. In the northeastern US, many prehistoric archaeological sites lack organic material suitable for direct dating using radiocarbon methods, largely due to very poor preservation in well-drained, acidic soils. Often this means that sites and components are dated solely on the basis of artifact (especially projectile point) styles. Projectile points are classified based on morphology (e.g., Justice 1987; Ritchie 1971), and the contexts from which they were recovered are then assigned the absolute dates that have been obtained for similar materials in the region. The resolution available with typological cross-dating is generally very broad and therefore not always adequate for sorting out remains into contemporary components or making comparisons among sites.

One often-voiced criticism of using projectile point typologies for dating occupations is that the various point types do not necessarily represent discrete temporal periods (Filios 1989), and that many of the types were likely used for extremely long (thousands of years) periods of time. In addition, more than one type may have been used by the same cultural group at any particular time (Snow 1980:162). Thus there may be some degree of temporal overlap among many of the described point types, even in a relatively restricted geographical area. Despite these and other drawbacks (e.g., Hoffman 1983), projectile point typologies are indispensable tools for Northeastern archaeologists (McBride and Dewar 1981; Starna 1979).

Ritchie's Typology and Nomenclature for New York State Projectile Points

Projectile points from northeast US archaeological sites are most often classified using the typology constructed by former New York State Archaeologist William Ritchie (1961, 1971), who during many years of research examined 10,800 artifacts from New York in addition to materials from a series of sites in coastal
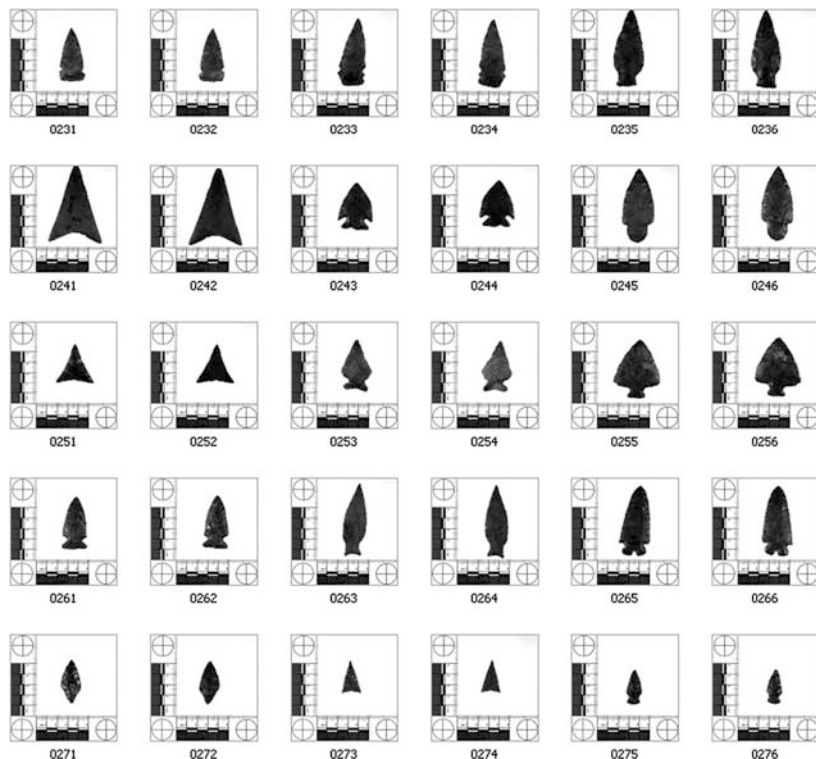
Fig. 13.1   A small portion of the over 10,800 projectile points analyzed by William A. Ritchie for his 1961 publication *A Typology and Nomenclature for New York Projectile Points*

Massachusetts (Fig. 13.1). Points were placed into types using chronological, geographical, and most importantly, morphological (size and shape) data. Ritchie first defined major categories based on seven shapes: side-notched, corner-notched, straight-stemmed, contracting stemmed, expanding stemmed, triangular, and lanceolate (Fig. 13.2). Next, these categories were further subdivided into types, often named for the site where they were first identified or recovered in high frequencies (e.g., Madison (Fig. 13.3), Levanna, Beekman, and Squibnocket triangles). In all, 27 projectile point types were described by Ritchie's original study (Ritchie 1961), and another 10 types were added for the revised edition published in 1971.

Problems with Ritchie's typology were evident from the start. More than 900 of the 10,800 points could not be typed. Further, the subjective nature of the classification scheme has been noted by several researchers. For example, three well-known and respected northeastern US archaeologists were independently asked to identify the projectile point types from one Hudson Valley, New York archaeological site. Although there was agreement among the three analysts for some of the artifacts, in many cases, two or three different types were assigned to a single piece (Claassen 1995). Another fundamental problem in dealing with artifact morphology concerns

Fig. 13.2 Primary morphological features presented by Ritchie (1961, 1971). Image reproduced with permission of New York State Museum
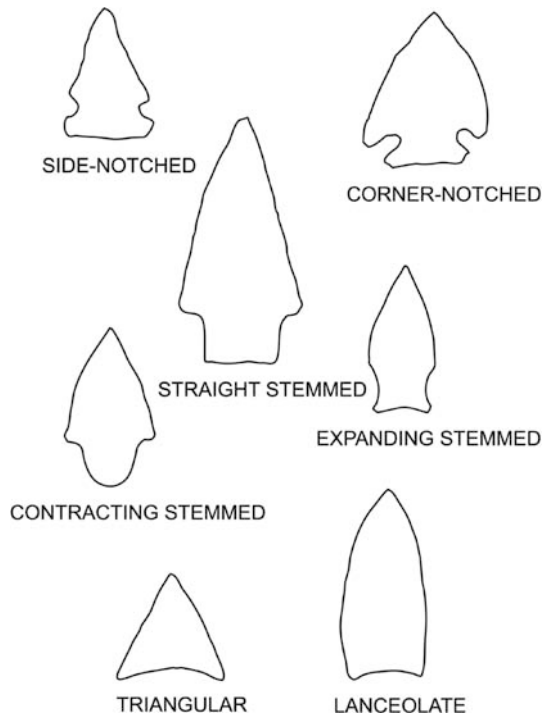
the use life of a stone tool. A projectile point is just one part of a continuum of stone tool technology that begins with raw material procurement and ends with eventual loss or discard, often with many stages of use and refurbishment along the way. Since making and recycling a stone tool is essentially a reductive process, a single piece may change shape as it diminishes in size over its use life (Frison 1968). This process, referred to as the "Frison Effect," can confound artifact typologies.

In this paper, we propose that alternative approaches afforded by computer vision and morphometrics can shed light on these and other problems of traditional stone tool typology. With the advent of computer vision, we can now examine the entire morphological continuum of projectile points through statistical shape analysis. Following automated image capture, three analytical methods were evaluated using silhouette, outline, and landmark data. Biased free capture methods to record the form of an individual projectile point, together with using invariant shape descriptors to quantify the data, may result in more objective analysis than was possible in the past.

Materials and Methods

A combination of custom and off-the shelf hardware and software was used to create a flexible image capture system (Fig. 13.4). The goal was to design a system

Fig. 13.3 Madison projectile points (Plate 16) in Ritchie's typology. As is evident in this example, a particular type as defined by Ritchie sometimes encompasses a high degree of shape variation (e.g., here these *triangular* points are lumped as a group despite having sides that are concave, convex, and straight). A number of hypotheses could potentially explain this problem: more than one point type is actually represented in this grouping, shape variations are due to unknown functional differences, or cultural reasons (e.g., importance of shape standardization). Image reproduced with permission of New York State Museum
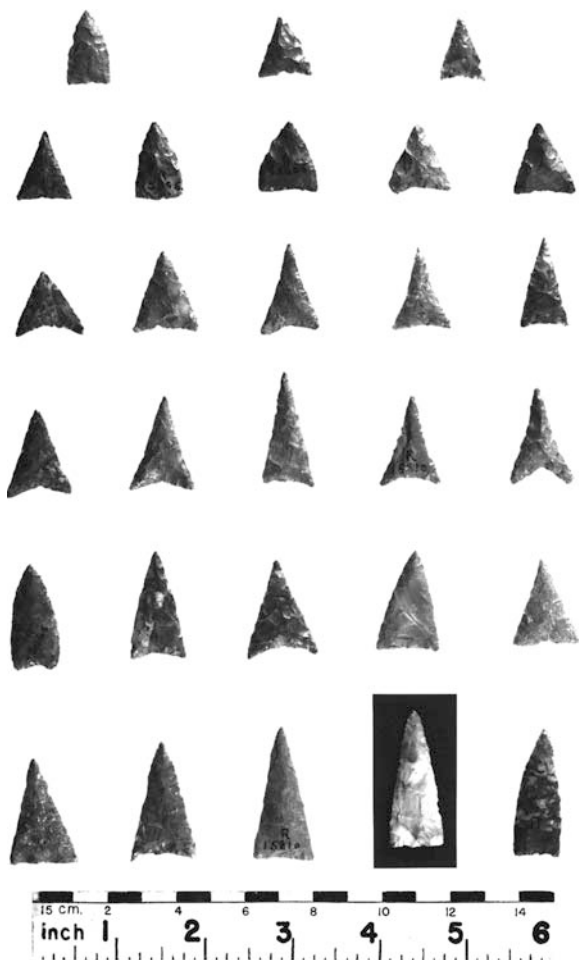
well-suited for fast, efficient data collection for statistical analysis of hundreds, if not thousands, of projectile points. The following five criteria were important for achieving this goal:

(1) the system should have a high speed image capture rate without compromising accuracy,
(2) be flexible in case any requirements changed,
(3) be modular, so it could be dismantled and transported to other research institutions,
(4) limit overall cost due to economic realities, and
(5) optimize the system workflow, in terms of image and data processing, to be as fully automated as possible without sacrificing any camera control, image
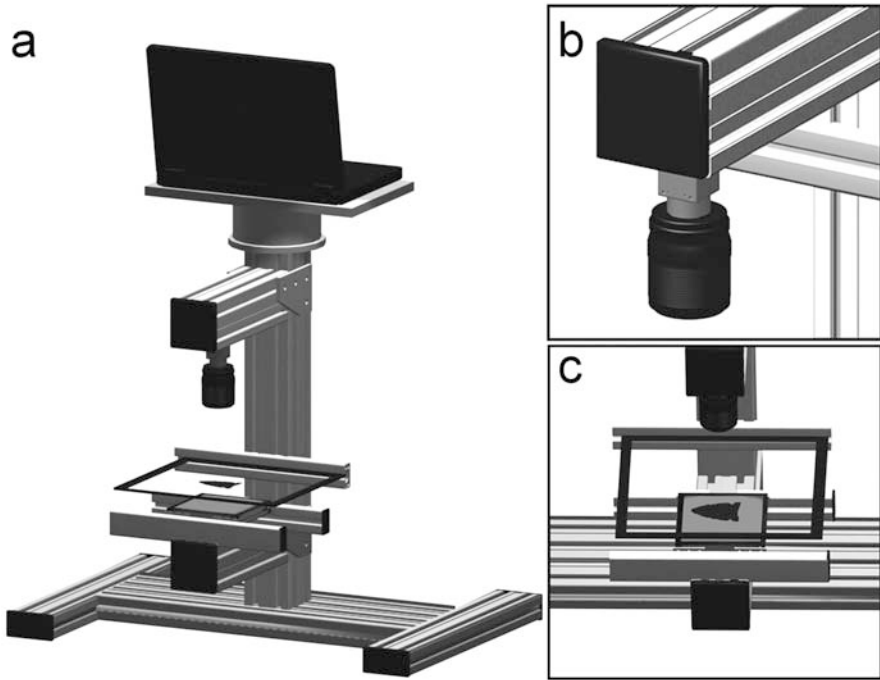
Fig. 13.4 Camera and laptop computer setup (**a**) used to capture projectile point images at high speeds without sacrificing shape or measurement accuracy. Detail of camera and telecentric lens with deeppass filter (**b**), and glare-free artifact stage, light-control filter, and LED backlight illuminator (**c**). Approximate capture rate is 70 projectile points per hour

processing capabilities or compromising the quality of data necessary for robust statistical shape analysis.

Hardware

Industrial structural aluminum framing was used as the main construction material for its rigidity and light weight (Fig. 13.4). The modular nature of the aluminum framing is ideally suited for the camera setup because it allows for the flexible mounting of equipment (e.g. lights, power boxes) and cable management. More importantly, precise adjustments required for different fixed and adjustable focal length lens and cameras can be made.

Three digital cameras are currently being evaluated for use with this setup: two Firewire (IEEE 1394a, DCAM 1.31 compliant) cameras, and an 8.0 megapixel Olympus digital SLR camera controlled by computer via tether. The choice of camera used for this kind of research is influenced by the method of imaging being

performed. For example, the setup described here consists of a camera, lens, filter and illumination optimized to capture accurate 2D shape silhouettes from 3D projectile points (see below).

A monochrome CCD-based Firewire camera was chosen for this study because it produces crisper edges than a comparable color camera (Steger et al. 2008). This is due to differences in the internal sensor array layout between monochrome and color charge-coupled device (CCD) sensor arrays. The color camera has an alternating array of red, green, and blue (RGB) light sensitive "wells" known as a Bayer pattern. Monochrome CCD sensors differ in that each well measures light intensity, not color. Silhouettes captured by monochrome CCD sensors have more edge fidelity than RGB color units (Russ 2002).

A telecentric lens was chosen for the camera because it eliminates common types of distortion that plague many compound lens systems such as parallax, pincushion, and barrel distortion. There are two caveats in using a telecentric lens effectively: 1.) the lens diameter has to be at least as large as the object being photographed, and 2.) a matching telecentric illumination source should be used in conjunction with this type of lens (Steger et al. 2008). Due to economic constraints, a partial telecentric lens was used (Computar Tec 55). The level of lens distortion was tested using a standard ISO 12233 slanted-edge measurement target and software designed to measure parameters such as the modulation transfer function (MTF). The MTF measures the spatial frequency response of the camera and lens system as a whole. The chosen lens was found to be sufficiently free of distortion for the purposes of this study.

The final elements of the system constitute the illumination system, probably the most difficult and counter-intuitive part of the entire system to design. The illumination required five specialized parts to work: an illumination source, a constant current source, a separate voltage supply, a specialized piece of light control film, and a specific lens filter mounted to the camera.

The illumination is provided by a red LED backlight (660 nm wavelength), driven by a constant current source and separate voltage supply to provide flicker-free operation. As stated above, monochrome cameras produce sharp silhouette images due to the internal layout of the CCD sensor array. The red color illumination was employed because CCD is most sensitive to this particular wavelength (Russ 2002). Telecentric lenses require telecentric illumination, which can be cost-prohibitive. Telecentric illumination consists of a large diameter compound lens which provides straight light rays in the same manner as collimated light sources used for microscopy. As mentioned earlier, the lens size has to be at least as large as the object being imaged, which equates to a telecentric illuminator costing many thousands of US dollars. Industrial engineers specializing in computer vision applications have developed an effective lower cost solution called light control films (personal communication, Daryl Martin, 2008). These plastic films have micro-louvers that block reflected light at known angles effectively creating a telecentric LED panel.

The last piece of the puzzle is the lens filter. Since it is desirable for the telecentric lens to capture only the straightened light produced by the red LED backlight, a deep red bandpass filter was used to eliminate all ambient light and just allow the 660 nm + 40 nm wavelengths through. According to one engineer (personal communication, Daryl Martin, 2008), this system may be capable of obtaining accuracies close to 1/1000 inch (0.025 mm) based on similar systems used in industry, though this has not yet been verified with the projectile point study.

Software

The software used for automated image acquisition and processing is VisiQuest from AccuSoft Corporation. VisiQuest employs a visual data flow programming paradigm that encapsulates blocks of code as "glyphs." The first glyph in the upper portion of Fig. 13.5 represents the image source (e.g., a camera) or from a directory of image files. The image data is then pipelined through a series of glyphs which represent a range of standard image processing (e.g., edge detection shown in the
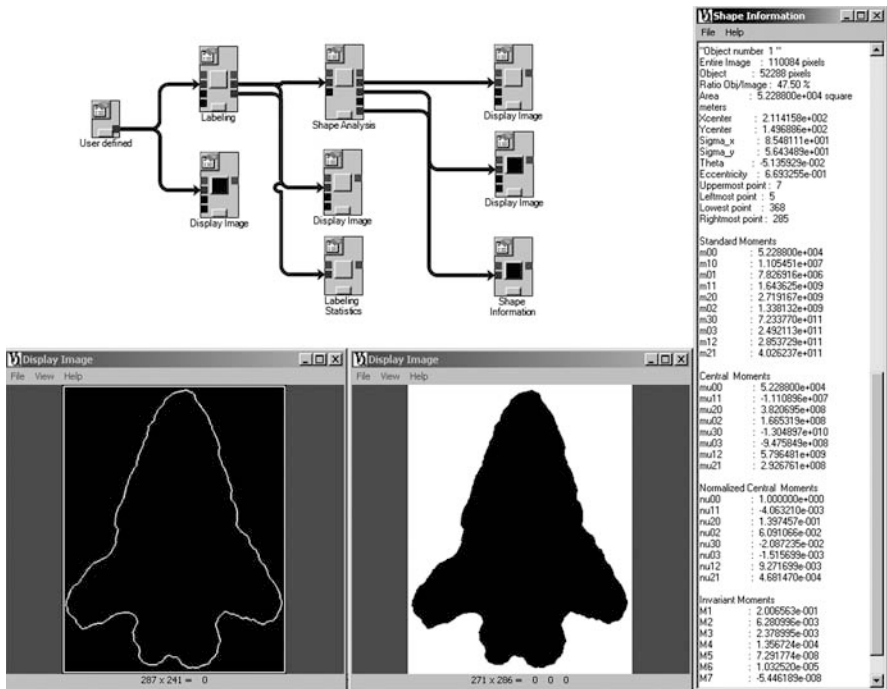
Fig. 13.5 Computer vision software (VisiQuest from AccuSoft Corporation)

lower left of Fig. 13.5, with corresponding statistical data to the right), morphological, and statistical tasks. The VisiQuest software is programmable with custom algorithms, but most applications can be found among the library of approximately 500 available functions.

Three approaches to statistical shape analysis were evaluated in this study: silhouette, outline, and landmark. The silhouette analysis was undertaken with a custom-written program developed by Holly Rushmeier and Andreas Glaser of the computer graphics group based in Yale University's Computer Science Department. Outline analysis was done using VisiQuest. Landmark data analysis was investigated using tpsDig2 (Rohlf 2008), tpsSmall (Rohlf 2003), MorphoJ (Klingenberg 2008), and PAST (Hammer 2008).

Analysis

As discussed above, among the goals of this pilot study was to evaluate various methodologies for analyzing data extracted from images in either silhouette, outline, or landmark form. Rather than immediately attempting to replicate Ritchie's typology study with 10,800 projectile points, we limited our sample size for each evaluation. This way, we hoped to limit any wasted effort in case one technique was deemed ineffective or found to be totally inappropriate.

The type of archaeological research proposed here has the potential to amass a very large amount of data, with sample sizes in the thousands, due to the high image capture rate possible using the camera setup described in detail above (with an average capture rate of 60 to 90 seconds per image). Further, this will be an ongoing multi-year data collection effort. The image capture camera system was designed to be disassembled for transport in order to image projectile point collections located at other museums, universities and archaeology laboratories.

Silhouette Based Methods

The first analytical method evaluated deals with silhouette data. Silhouettes were derived from raw image data by threshold and segmentation algorithms to create black and white bitmap images of each side of 100 specimens. The bitmap data were then subjected to geometric harmonic analysis that uses diffusion distances to organize data (similar to Isomaps used in cluster analysis). These methods create a shape space defined by the group of objects themselves. The original bitmaps were scaled and plotted as data points in such a shape space (Fig. 13.6). We inadvertently included images of both sides of each projectile points, which explains the pairings evident in Fig. 13.6. This actually may provide a useful statistic for ascertaining and quantifying effects of symmetry/asymmetry within a sample (as defined by a vector between specimen centroid pairs).

Fig. 13.6 Organization plot
based on natural basis
functions. This plot was
generated by custom
programming code written by
Dr. Holly Rushmeier and
Andreas Glaser of the
computer graphics group
based at Yale University

Outline Based Method

After an object has been imaged and converted to a bitmap silhouette, an accurate
outline may be obtained by using a variety of edge detection algorithms com-
monly used in the image processing field. The outline can then be converted to
a series of x-y coordinates by placing a grid over the outline and "cutting" out
points. The x-y coordinate pairs then are averaged to determine the centroid of
the outline. If measurements are taken from the centroid to the object outline at
known angles (polar coordinates), a signature plot of the object can be created
(Gonzalez et al. 2004). While this method is useful for a variety of tasks (e.g.,
Optical Character Recognition and pattern/template matching), by itself it is not a
robust statistical shape analysis technique for our research needs because it fails with
certain projectile point shapes (personal communication, Holly Rushmeier, 2006).
For example, taking measurements from the centroid of a corner-notched projectile
point (Fig. 13.2) will result in more than one intercept on the outline near the point's
base from a given polar coordinate, invalidating the data.

Signature plots may still prove useful for mathematically determining geomet-
ric landmark data. This is accomplished by using a so-called "watershedding"
algorithm, in effect filling the troughs of the signature plot graph, whereby the min-
ima/maxima points can be used to place objective (mathematical) landmarks on an
amorphous set of shapes, thus overcoming the problems presented by irregularly
shaped projectile points (Fig. 13.7).

Fig. 13.7 Signature plots of three projectile points representing the outline of each specimen. Landmarks may be determined using a watershed algorithm to find minima/maxima locations on the plots
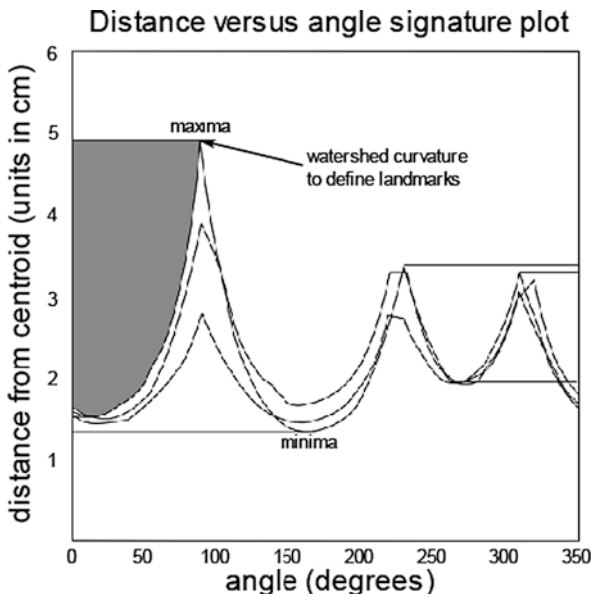
Landmark Based Methods

A small sample (n=33) of triangular projectile points was digitized manually using tpsDig2 (Rohlf 2008). The points were previously classified by Ritchie as Levanna and Madison (Fig. 13.3) types. In this pilot study, landmarks were not rigorously determined, but rather they were subjectively placed in an attempt to describe features normally noted by archaeologists (e.g., height relative to width, curvature of the sides). Six landmarks were located on each of the 33 specimens. The first (landmark 1) was placed at the presumed top of the projectile point. The remaining five landmarks were then recorded counterclockwise from the first: two were placed at the remaining corners (landmarks 3 and 5), and three were placed at the midpoints along each side (landmarks 2, 4 and 6) to reflect the degree of concavity or convexity of each side (Fig. 13.8). Principal component analysis was performed on the data (Fig. 13.9; Tables 13.1, and 13.2), followed by generalized Procrustes analysis (GPA). While the choice of landmark location and number could be made more objectively in future research, the six landmarks chosen here were sufficient to differentiate two distinct projectile point types in a robust fashion (Figs. 13.10, and 13.11).

Cluster analysis was performed on the small sample of projectile points (n=33) using PAST software (Hammer 2008). The tree was created using the unweighted pair-group method using arithmetic averages method UPGMA (Sneath and Sokal 1973) with the similarity measure set to use Euclidean distances. The cophenetic correlation coefficient was 0.9132. Four of the points (specimens 0, 30, 31,
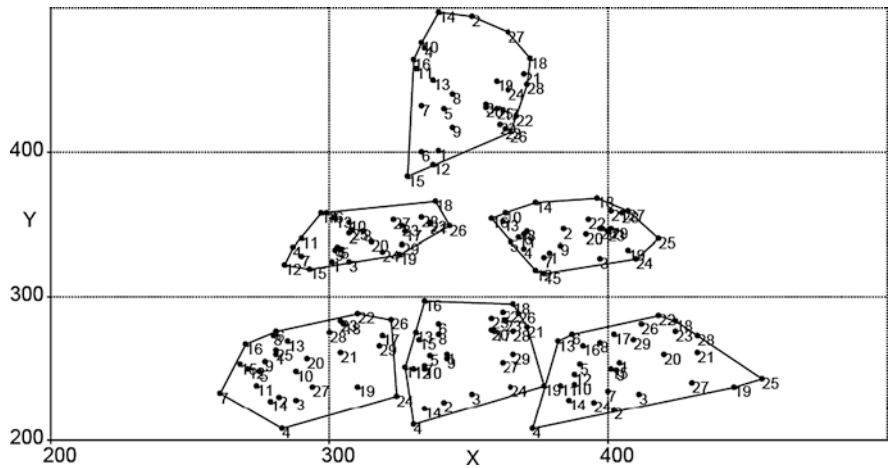
Fig. 13.8 Landmark data before being subjected to generalized Procrustesanalysis (GPA). As this was a pilot study, only six landmarks were used (three at each corner and three at the midpoint of each side) on a small sample (n=33) of triangular projectile points
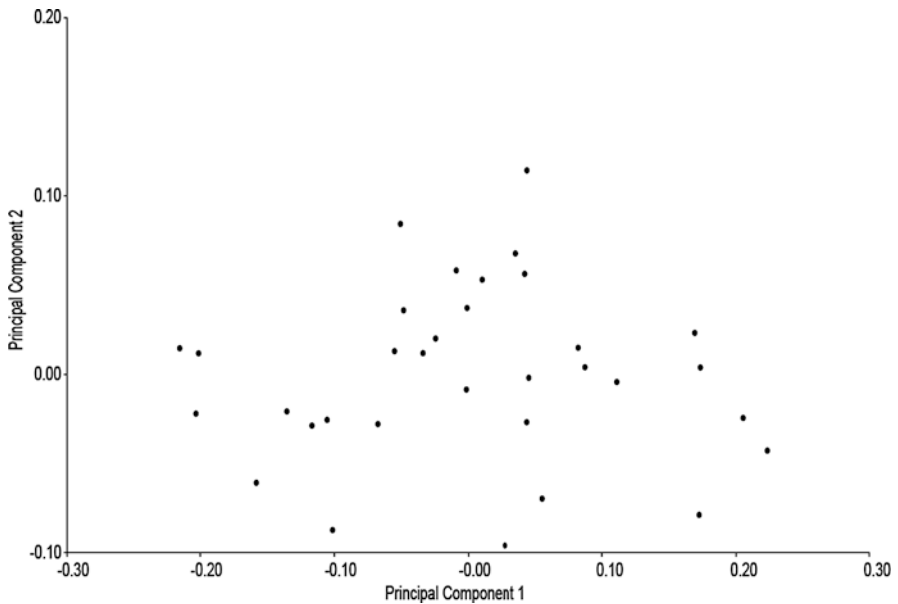
Fig. 13.9 Principal component analysis of 33 triangular shaped projectile points. PC1 (73.55%) and PC2 (13.90%) account for 87.45% of the shape variation within an unclassified sample of Madison and Levanna points. See Tables 13.1, and 13.2 below

Table 13.1 Eigenvalues. Principal components 1 through 8

Eigenvalues	% Variance	Cumulative %	
1	0.01137	73.547	73.547
2	0.00215	13.899	87.446
3	0.00099	6.425	93.871
4	0.00035	2.266	96.137
5	0.00028	1.807	97.944
6	0.00013	0.84	98.784
7	0.00012	0.802	99.586
8	6.4E-05	0.414	100

Table 13.2 Principal components analysis

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
x1	0.02294	0.03831	−0.0993	−0.3215	0.40159	0.01424	−0.3508	0.1365
y1	0.40364	0.1915	0.09103	0.1826	0.29034	−0.2418	0.18443	0.01971
x2	−0.2454	−0.5554	0.33308	−0.0936	−0.1464	−0.0605	0.50789	0.05397
y2	0.06558	−0.168	−0.3092	−0.522	−0.4613	−0.1621	−0.2282	0.28729
x3	−0.558	0.23608	−0.2143	0.29778	0.05767	0.18084	−0.1595	0.27183
y3	−0.1163	−0.1473	−0.267	−0.2313	0.57376	0.17295	0.3238	−0.1146
x4	−0.0044	−0.0047	0.08788	−0.0914	−0.0171	−0.3468	−0.2996	−0.7264
y4	−0.2469	0.43693	0.66814	−0.1471	−0.0598	−0.0761	−0.0406	0.15251
x5	0.55272	−0.2209	0.20057	0.25025	0.04502	0.13551	−0.1883	0.35405
y5	−0.1558	−0.1902	−0.253	0.56852	−0.1132	−0.4171	−0.0617	0.0073
x6	0.23213	0.50666	−0.308	−0.0415	−0.3408	0.07678	0.49026	−0.0899
y6	0.04977	−0.1229	0.06999	0.14932	−0.2298	0.72419	−0.1777	−0.3522

and 32) clustered on the branch farthest left had all been previously classified as Levanna points by Ritchie (1961, 1971). One previously identified Levanna point (specimen 1) was reclassified as a Madison point by the cluster analysis.

Discussion and Conclusions

Traditional artifact typology, although an indispensable tool for archaeologists and potentially imbued with cultural meaning, is limited by problems as illustrated with Ritchie's (1961, 1971) classification system for New York State projectile points. The bias free data capture methods made possible by computer vision, together with statistical methods that analyze shape in terms of invariant descriptors, may offer some solutions. Other archaeologists using this or similar approaches include Bradbury and Richmond (2004), Buchanan (2005), Cardillo (2006), Crompton (2008), Gero and Mazzullo (1984), Lohse et al. (2004), Pinkerton (1979), Saragusti et al. (2005), and Wei et al. (2007).

A computer based system, which relies on clearly defined geometric rules and algorithms, has the potential to limit the amount of subjective bias in classifying
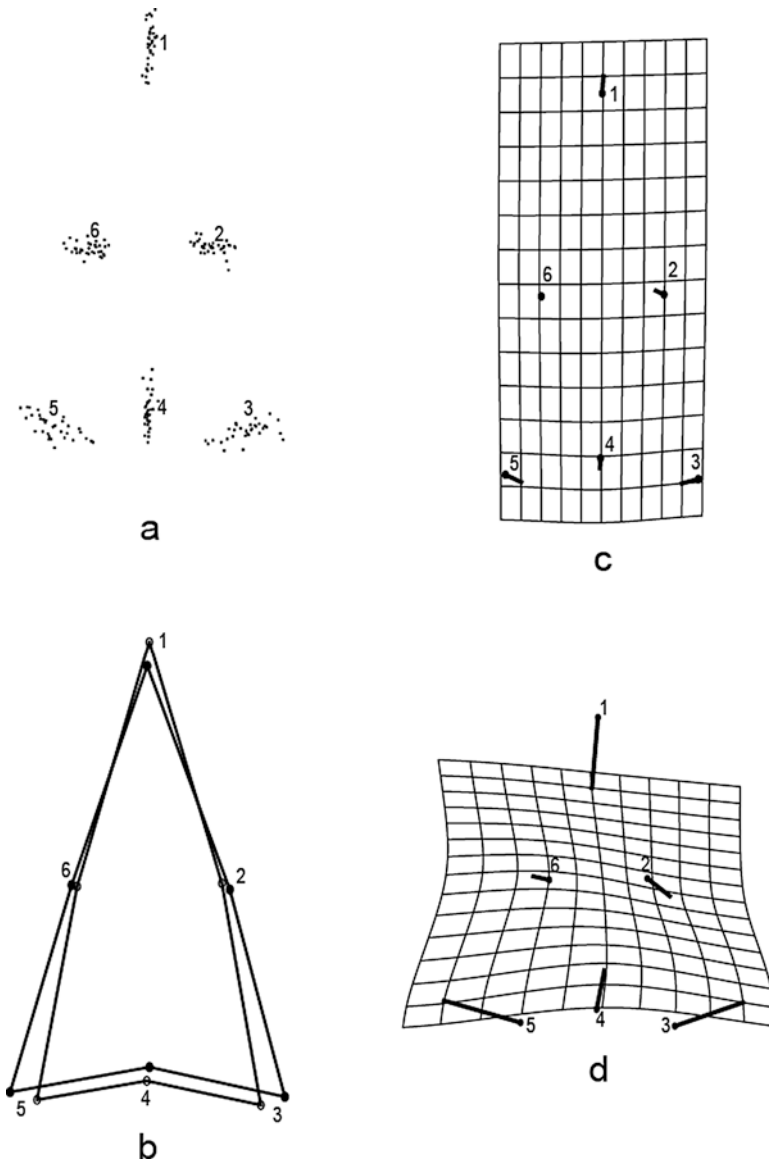
Fig. 13.10 The six landmarks shown clustered after generalized Procrustes analysis (**a**), wireframe connecting landmarks of PC1 (narrow *triangle*) over average shape of the sample (**b**), transformational grids (**c** and **d**) showing variation within two classified groups: Madison (**c**) and Levanna points (**d**). Note that the larger variation in Levanna points (**d**) is most likely due to the small sample size (n=5)
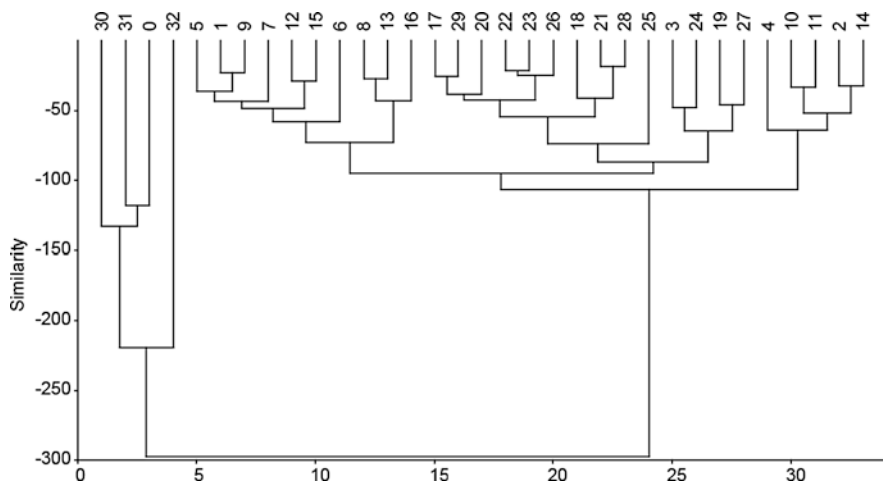
Fig. 13.11   Cluster analysis of 33 triangular projectile points

artifacts such as projectile points. Another benefit of using computer vision is fast, efficient, and accurate data capture, and is thus well-suited to documenting large artifact assemblages. We describe three analytical methods (silhouette, outline, and landmarks) that may be used to quantify artifact shape variables, and allow us to compare projectile points with one another in the context of a shape continuum as well as a discrete point type.

Research Potential of Museum Collections

It is generally recognized that the majority of museums and other data repositories often exhibit only a fraction of their collections. Especially in the case of material recovered from an archaeological site, once an artifact has been cleaned, cataloged, and curated, it may only infrequently be re-examined, if at all. This has significant consequences for contributing to our knowledge and understanding of the past (Cantwell et al. 1981).

Perhaps among the most valuable products of the high speed data capture method described in this paper is that it will result in a digital photographic record of large collections of projectile points which could potentially be repurposed to answer a range of research questions. The ability to visually query a database, as well as extract useful shape and metric data directly from images of artifacts, has significant research potential and will likely lead to new perspectives on common archaeological problems. For example, by recasting the projectile point from a discrete and fixed "type" to instead as one stage in a continuum (from raw material procurement, through manufacture, use, resharpening, to loss or discard), we may be able to reconstruct a fuller view of ancient human behavior. Projectile

points may also be considered as part of a continuum of shape variation that changes in multiscalar dimensions due to a variety of mechanisms such as cultural data transmission (O'Brien 2008), need-based functional changes in morphology (e.g., changes in prey animals and hunting strategies), or simply the summation of divergent variations and characteristics which accumulated over time as groups of human populations spread out over the landscape, divided by natural physiographic or invented cultural boundaries. Linking artifact shape data with Geographical Information System (GIS) datasets is likely to reveal how the sources of variation within the morphometric data are distributed in both spatial and temporal dimensions.

References

Bradbury AP, Richmond MD (2004) A preliminary examination of quantitative methods for classifying small triangular points from late prehistoric sites: a case study from the Ohio River Valley. Midcontinental Journal of Archaeology 29:43–61

Buchanan B (2005) Cultural transmission and stone tools: a study of Early Paleoindian technology in North America. Ph.D. dissertation, University of New Mexico, Albuquerque

Cantwell AM, Griffin JB, Rothschild NA (1981) The research potential of anthropological museum collections. Annals of the New York Academy of Sciences Volume 376. New York

Cardillo M (2006) Temporal trends in the morphometric variation of the lithic projectile points during the Middle Holocene of southern Andes (Puna Region): a coevolutionary approach. Paper presented at the XV World Congress of the International Union for Prehistoric and Protohistoric Sciences. Lisbon, Portugal

Claassen C (1995) Dogan point: a shell matrix site in the lower Hudson Valley. Occasional Publications in Northeastern Anthropology, No. 14. Archaeological Services, Bethlehem

Crompton S (2008) 3D lithic analysis. In: Posluschny, A, Lambers, K, Herzog, I (eds.) Layers of perception: proceedings of the 35th international conference on computer applications and quantitative methods in archaeology. Dr. Rudolf Habel, Bonn, Germany

Filios EL (1989) The end of the beginning or the beginning of the end: the third millennium B.P. in southern New England. Man in the Northeast 38:79–93

Frison GC (1968) Functional analysis of certain chipped stone tools. American Antiquity 33: 146–155

Gero J, Mazzullo J (1984) Analysis of artifact shape using Fournier series in closed form. Journal of Field Archaeology 11:315–322

Gonzalez RC, Woods RE, Eddins SL (2004) Digital image processing using MATLAB. Pearson Prentice Hall, Upper Saddle River

Hammer Ø (2008) PAlaeontological STatistics (PAST), version 1.89. Paleontological Museum, University of Oslo, Norway. http://folk.uio.no/ohammer/past/

Hoffman C (1983) Radiocarbon and reality: the fifth millennium BP in southern New England. Man in the Northeast 26:33–53

Justice ND (1987) Stone age spear and arrow points of the midcontinental and eastern United States. Indiana University Press, Bloomington

Klingenberg CP (2008) MorphoJ. Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm

Lohse ES, Schou C, Strickland A, Sammons D, Schlader R (2004) Automated classification of stone projectile points in a neural network. In: Enter the past. The e-way into the four dimensions of cultural heritage. Computer applications and quantitative methods in archaeology. BAR international series 1227, Archaeopress, Oxford

McBride KA, Dewar RE (1981) Prehistoric settlement in the lower Connecticut river valley. Man in the Northeast 22:22–37

O'Brien MJ (2008) Cultural transmission and archaeology: issues and case studies. Society for American Archaeology, Washington, DC

Pinkerton KG (1979) Multivariate morphometrics applied to projectile points. M.A. thesis, Colorado State University, Fort Collins

Ritchie WA (1961) A typology and nomenclature for New York projectile points. New York State Museum and Science Service bulletin 384. State University of New York, Albany

Ritchie WA (1971) New York projectile points: a typology and nomenclature. Revised edition. New York State Museum and Science Service bulletin 384. State University of New York, Albany

Rohlf FJ (2003) tpsSmall, version1.20. Department of Ecology and Evolution, State University of New York, Stony Brook

Rohlf FJ (2008) tpsDIG2, version 2.12. Department of Ecology and Evolution, State University of New York, Stony Brook

Russ JC (2002) The image processing handbook. Fourth edition. CRC Press, Boca Raton

Saragusti I, Karasik A, Sharon I, Smilansky U (2005) Quantitative analysis of shape attributes based on contours and section profiles in artifact analysis. Journal of Archaeological Science 32:841–853

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Snow DR (1980) The archaeology of New England. Academic Press, New York

Starna WA (1979) Late Archaic chronology for the Middle Mohawk Valley, New York State: a review of the type concept and cross-dating. Man in the Northeast 17:3–18

Steger C, Ulrich M, Wiedermann C (2008) Machine vision algorithms and applications. Wiley-VCH, Weinheim, Germany

Wei L, Keogh E, Xi X, Lee S-H (2007) Supporting anthropological research with efficient rotation invariant shape similarity measurement. Journal of the Royal Society Interface 4:207–222

Chapter 14
Morphometric Analysis Applied to the Archaeological Pottery of the Valley of Guadalquivir

Ana L. Martínez-Carrillo, Manuel Jesús Lucena, José Manuel Fuertes, and Arturo Ruiz

Idea and Aims

Ceramics are one of the most documented materials in the archaeological interventions. The documentation and the analysis of the pottery shapes allow the knowledge of the chronology and the functionality of the settlement where they have been found.

The achievement of a typology of ceramic materials is made attending on different aspects (function, context, morphometry . . .).

In this contribution a methodology of analysis of archaeological ceramic is showed. This methodoly is based on the technique of nonrigid deformable analysis applied to the drawing of the profile and is aimed at the construction of a ceramic typology.

Introduction

The pottery forms are one of the most documented in the archaeological interventions, and therefore is the most voluminous, allowing greater information that contributes to the reconstruction of the historical sequence of a place (Orton et al. 1993). The study and analysis of the ceramics constitutes one of the most frequent activities of the archaeological work, which consists habitually of classifying the thousands of ceramic fragments gathered in the interventions and selecting those that contribute to deduce forms, functions and chronology.

The ceramic materials usually appear in fragmented state. In fact, except for the case of a sudden desertion of a place, with good conditions of conservation throughout the time or closed sets like graves, the most normal is to find a great amount of ceramic fragments.

———————————

A.L. Martínez-Carrillo (✉)
Centro Andaluz de Arqueología Ibérica, Universidad de Jaén, España, Spain
e-mail: anamartinezcarrillo@gmail.com

The study of archaeological ceramics means giving continuity to an archaeological investigation that has tried diverse classification approaches since the very origins of the discipline, and none of them can be considered valid.

The different criteria used in the elaboration of typologies do not contribute to homogenize the analysis of the pottery shapes, since the election of criteria depends on each researcher. The most used criteria have been the morphometrics, functional, statistical, technological and contextual. Taking this division into account, the criteria used in this research is morphometrics because, within this scope, methods and techniques of analysis coming from graphical computer science can be applied and can be useful for classifying pottery shapes according to degrees of similarity.

In this chapter a deformable contour based, computational method of comparison of profiles is exposed. This method allows the automatic execution of pottery shapes based on morphometrics criteria. This methodology of comparison has been practised on a sample of drawings of complete ceramic vessels and fragments, allowing not only grouping similar shapes but also assigning pottery fragments of the base or the rim to complete forms.

The present chapter has been structured as it follows: definition and evolution of the concept of *type* in the classification of ceramics, description of the area of study and the analyzed data, methodology of analysis and the conclusions.

The Classification of Archaeological Pottery: Conceptual Trends and Classifications of the Iberian Pottery

The Concept of Type

The increasing number of excavations since the nineteenth century and the need to establish typologies related to stratigraphical chronologies helped to the establishment of descriptions, typological sequences and determination of *fossil-guides*.

The multiplicity and diversity of criteria regarding the ceramic classification have produced, since the second half of the twentieth century, *typological* discussion, put forward by authors such as Adams (1988, pp. 44–42) or Orton et al. (1993, pp. 21–26).

Generally the objectives of a classification are the arrangement and cataloguing of the material, the relative dating of the archaeological contexts and the establishment of parallelisms with materials documented in other zones. The accomplishment of a typology is an empirical practice, oriented to facilitate the interpretation, but isolated and without theoretical expositions in which to sustain itself.

At the end of the 1960s an important advance of the General Theory of Archaeology was achieved with the publication of the work of D.L. Clarke *Analytical Archaeology*. From this moment on, the use of mathematical methods applied to Archaeology commences with the purpose of being able to make the data more objective. One of the added concepts in this book is the concept of *type*, which is considered as *the element that is related to a more or less ample group*

through a series of common attributes. An exact place and a chronological context differentiate it from other types (Clarke 1984).

Therefore, in the establishment of a typology is necessary to know clearly the concept of *type*, since its meaning varies in the space and in the time. Every type has to show two basic properties: identity and meaning. In addition, a type is defined in relation to a specific classification that partially indicates the rules by which that is formulated. Furthermore, a typology is a system and not only a collection of types, it must be coherent as a system.

Therefore, according to these postulates, in the elaboration of a typology the types are defined first, and then subtypes or variants are established. The differences between types are marked by the presence or absence of the most significant atributes.

This concept of type and its supposed implications were the basis of the difusionism that dominated the prehistoric archaeology of the 1950s and the 1960s.

From the seventies these concepts are reviewed, generating a wide discussion. The discussions is considered from the problems derived of the meaning that was attributed to the functional identity of a vessel, since that characteristic frequently began to be used for the identification of the archaeological contexts. The situation arrived at a point in which the chronology, the functionality of the establishments or the cultural identification was even established on the basis of the percentage of different vessels entered in the site. As Shennan indicates, the quantitative information is an essential part of the archaeological work, in this sense the use of the computers is a useful tool for the data processing (Shennan 1992).

The methodological development of the Archaeology during the last thirty years has allowed the incorporation of quantification techniques in the archaeological studies. For this reason, against the logic of the method, the critics argued that the used criteria were intuitive and subjective, since was the researcher that established the most significant characteristics; reason why the type is defined only represented an imaginary ideal.

In the last years has rejected practically the accomplishment of a ceramic classification considering a single approach of analysis, because taking conclusions from a single element leads to an erroneous classification, since the presence of determined material can be due to many different factors. It is obvious that in many occasions the classification or the types established with the interpretation of the context is being confused. Contextual archaeology includes diverse scales and space, hierarchic and ecological dimensions. This approach can be applied to simple or complex societies (Butzer 2007).Therefore the context in which an artefact is found has special interest for the later analysis of the site.

Ceramic Typologies for the Iberian Period

One of the major tasks of the archaeologists in relation to the ceramics has been the creation of ceramic types able to respond to the different questions raised by the

investigation. The study of ceramic forms can focus from the point of view of the functionality, aesthetic or taxonomic (Shepard 1956).

The function makes reference to the human interests; the intention of a container says to something about the activities and the customs to us of the people used who it. Despite there is no an exact correspondence between the form and the function, since a ceramic form can have several functionalities and functionality can be represented by different forms.

Like main function, the ceramic containers are made for storage, processing of raw materials and transport. The functionality of the ceramic containers has been treated in different works, between which it is possible to mention (Braun 1983; Oswalt 1973; Lustic-Arecco 1975).

As far as the aesthetic aspects, trims in the contour and the proportion of a container, less have been considered in the field of archaeology. There is a direct relation between the development of a style and the rank of detail of his attributes.

In the construction of a taxonomy different nomenclatures have been used to define the ceramic shapes. This method is useful to order a certain ceramic sample; although he is not very effective to make studies comparative of different places.

The data of the ceramic forms of different stages and regions have been accumulated until the point of which an efficient system of storage of images is required. The necessity of a classification method exists that includes and stores the attributes of all the ceramic forms.

Before this theoretical emptiness the revision becomes necessary of the different works for the improvement from the own practice, from new technical incorporations yielded by other sciences that concur on the material object of the typology. In a parallel way is also made a theoretical exposition necessary that it leads to a methodologic elaboration that fixes when, how and mainly why of the own typological practice (Ruiz and Molinos 1993).

In the last years the archaeologists are shortage computer science and the statistic; lamentably these discoveries have been applied with enough delay to the Iberian culture, and have used who have dedicated them their works to analysis of sets of graves, or to the closed spaces of the houses of a town, but not to the typologies.

The statistics is valid by its capacity to order the appearances of the observed evidences. On another hand, is not more than the contrast of the description of an apparent phenomenon with a model of probabilistic operation. Really, it allows verifying us but without contrasting due to the capacity to a more objective discovering.

There is not to forget that this computer science-statistical project constitutes only one of the new possibilities of coordination of practices with objectivity aims. That is to say, the new articulation allows replacing the formal analysis by the morphometric analysis. Likewise, the practices of chemical analyses and physical open a new way to technical variables that previously had been only evaluated subjectively.

If the structuralism or the neofunctionalism feeds a new reframing on the conception of object, and therefore, of their variables in the new hierarchic frame that supposes the classic historical-space-temporary classification. This also affects the concept of cuts stratigraphic provider of the typological material, since it is due to

interpret that a sequential cut indiscriminately crosses several spaces with different functions. Thus the possibilities of resisting the evolution or the tendency of the types decrease considerably from an objective point of view.

Consequently structural equality is due to demand, at least, for the sequential or even macrospace contrast (different sequences from different settlements) from the functional-space point of view in all the cases. It does that, so that the typology in a frame superior or different from which is exclusive and own of the arrangement, to make their chronological function really effective, is advisable to evaluate the methodological process that leads to select to a certain technique of excavation (Ruiz and Molinos 1993).

The morphologic examination of the archaeological ceramics can be expressed of different ways:

- A morphologic examination that considers the shape and the elements (rims, handles, bases).
- A more rigorous analysis than considers the structure of the shape, also the physical-chemical characteristics of the clay.

In this intention a disparity of methods, opinions and nomenclatures exists, being a great variety of analysis that varies with each author. However, a recent interest in integrating the ceramics in an ampler analysis of the sets can be warned. This approach depends on the nature of the findings and the way in which they have been registered. It is necessary to begin to integrate the analyses of ceramics within ampler scopes, that is to say, in the global sets of findings.

The archaeological investigation in the elaboration of typologies of Iberian ceramics has defined two different lines of work: synchronous and another diachronic one. First it is ligature to the cataloguing of sets of materials, in many cases without allegiance to stratigraphy. One has been used what could be described like synchronous models, that they only try to order in groups or types a mass of material and in any case and later to add when it is possible the chronological factor.

The second model, the diachronic one, tries to value the evolution of certain types. In the first case, the synchronous work is possible to emphasize the typology of Pla and Aranegui (1981) on the Iberian set of vessels, the one of Pereira (1979) on the materials of Toya, or analysis on very concrete types like the stamped ceramics. Within model diachronic, that usually is very present in the analyses of concrete settlements where the variable time does not display doubts, would be possible to mention realised by Ruiz Mata about La Torre de Doña Blanca (Ruiz Mata 1987), those of Pellicer on Cerro Macareno (Pellicer 1982) or those of Ruiz Rodriguez and others on the Cerro de la Coronilla (Ruiz et al. 1983).

In a second level of the typological investigation it is necessary to value the scope of the sample, that is to say, if one is a sequence exclusive of a settlement, of a region or zone or delimited, or in case on the contrary, one is a general typology of the Iberian period, model del that does not appear examples by the complexity that the case can raise. Therefore, throughout the archaeological investigation there has been no a criterion determined for the elaboration of a typology, because each author

applies the most advisable criteria to a determined group of forms. As departure point has been considered that a typology must be realised having in account as much specific criteria of the establishment like criteria of a level more general than they can be applied to any group of materials.

Inside the specific criteria of the establishment it is necessary to take care of the context of the finding mainly, because this context is going to determine many of the questions.

Area of Study and Documentation of the Material

The selected ceramic material for the morphometric analysis comes from different archaeological settlements located in the provinces from Jaén, Granada and Cordoba. Most of the ceramic vessels have been documented in the province of Jaén, since they pertaining to Iberian period. In this area has been one expanded tradition with respect to the study of ceramic typologies of the Iberian period, emphasizing the works of Pereira for the Iberian ceramics painted of the upon valley of Guadalquivir (Pereira 1989) (Fig. 14.1) (Table 14.1).
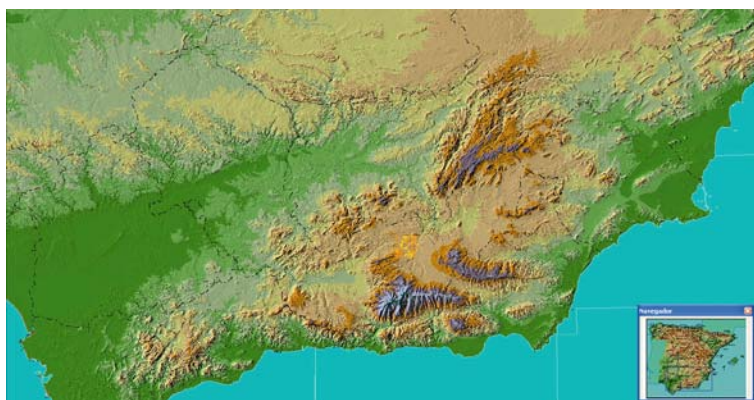
Fig. 14.1 Studied area

Table 14.1 N^o settlements, chronology and number of studied vessels

N^o settlements	Chronology	N^o vessels
2	VII BC-V BC	90
9	IV BC-III BC	829
4	II BC-II AC	255
1	III AC	85
Total		1259

The combination of different archaeological sites, with different chronologies makes the accomplishment of a diachronic and synchronous study possible, allowing to contrast materials of different archaeological sites with different chronologies.

The sample for the analysis has been made of 1259 complete forms whose chronology goes from the VII B.C to the S. II A.C., belonging to 16 archaeological settlements of the area of the Andalusia.

Documentation and Digitalization of the Ceramic Material

The documentation and traditional reconstruction of archaeological ceramics are based on the profile of the vessel that is the section of the fragment that contains the rotation axis. In the case of the analyzed ceramics is wheel made and is therefore symmetrical. The extraction of the profile is fundamental to be able to establish a classification according to the form of the same.

The process of digitalization of the images of the ceramic drawings can be divided in the following steps (Melero et al. 2003) (Fig. 14.2):

- Digitalization of the drawings of publications.
- Vectorization of the contours.
- Separation of the profiles for its later computer processing.
- Export of the drawings of the profiles to a format to raster without compression, in this case the test has been realized with PNG file formats.

The sample of the reference collection is formed by drawings of complete vessels, understanding as such drawings of complete profiles or fragment drawings had enough form information that is possible to reconstruct the complete section of the vessel.

The drawings of the complete vessels come mainly from publications. The documentation available has been compiled to homogenize the graphical and documentary information, which is not standardized and it does not follow canons at the time of his study and publication.

The computerization of the archaeological registry allows to homogenize the data that are introduced and to make agile the process of obtaining of results, reason why allows a standardization of the used methodology to compare results.
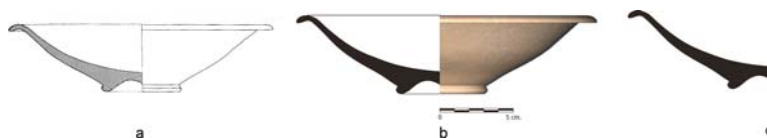
Fig. 14.2 (**a**) Digitalized drawing from a publication; (**b**) Previous image vectorization; (**c**) Vessel's profile

The representation of the drawing has normally been realized the representation of the profile usually imagines in left half of the drawing along with the internal decoration, whereas the other half is reserved for the outer representation of the vessel.

Once compiled all the publications in which they appear drawings of ceramic vessels it carries out a task of digitalization of these drawings to homogenize the visual reconnaissance of all the drawings and to vectorize, and therefore to even compress the space of each image for its later computer processing.

The digitalization process consists of scanning and vectorizing each one of the compiled ceramic forms.

Profile Comparison Method

In order to estimate the similarity between two profiles, a comparison technique based on non-rigid deformation analysis will be used (Fig. 14.3).

Fig. 14.3 Example profile image

Deformable Models

First of all, a measure to evaluate the *effort* or *deformation energy* needed to apply to a given contour in order to adapt it to another, is defined.

The deformable model given by Nastar (1994) and Nastar and Ayache (1996) is used, which is described briefly below.

This model was first used for analysing the non-rigid motion of structures in temporal sequences of 2D and 3D biomedical images.

The mechanical formulation of the problem consists in assuming that the contour is made up of a set of points (or nodes) with mass, joined together by springs. These elastic springs provide a *polygonal approach* of the contour and are supposedly identical, without mass, with stiffness k and length l_0. These springs modelize the *surface elasticity* of the object.

Therefore the system under study is composed of N points located at time t on geometric points $(M_1(t), M_2(t),..., M_N(t))$.

For all M_i, the basic equation of the dynamic expresses the vectorial sum of the external forces acting on M_i as equal to the product of its mass and its acceleration:

$$\sum_j \vec{f}_j(M_i,t) = m_i \frac{d^2}{dt^2}\overrightarrow{AM}_i = m_i \ddot{M}_i \qquad (14.1)$$

This equation, expressed for nodes N, leads to a *non–linear system of coupled differential equations* (for each node, displacements *x, y* and *z* are paired and the displacement of a node depends on the displacement of its neighbouring modes).

A possible solution to the non-linear equation given in (14.1) could be the linearisation described in Pentland and Sclaroff (1991). Nastar (1994) proposes making $l_0 = 0$.

The direct consequence of the assumption of the above hypothesis is that a linear system of differential equations is obtained, in which calculations can be separated into *x, y* and *z*.

Lastly, the deformation of the system is given in the form of a matrix of $3N$ dimension by:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\ddot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{F}(t) \qquad (14.2)$$

where $\mathbf{U}$ is a vector which stores the displacement of nodes, $\mathbf{K}$ is the system's stiffness matrix, $\mathbf{C} = \text{diagonal}\,(c_i)$ is the buffer matrix, $\mathbf{M} = \text{diagonal}\,(m_i)$ is the diagonal masses matrix and $\mathbf{F}$ is the force image which contains the object attracted by the contour images. Matrices $\mathbf{C, M}$ and, in particular, stiffness $\mathbf{K}$ matrix, are constant, which is why they should not be recalculated in each iteration. Equation (14.2) gives the formulation based on the use of finite elements of the deformation process.

In a dimensional space of order 3, the matrix equation (14.2) represents fitted differential equations $3N$. According to the hypothesis already made of making $l_0 = 0$, the result is an expression of the same given by 3 matrix equations of the N order for *x, y* and *z*.

These differential equations of motion can now be integrated according to the shape of their matrix following different methods (Bathe 1982; Nastar 1994) uses an explicit Euler scheme.

Modal Analysis

To provide a solution to the balance equation (14.2), the problem is transformed by changing the base: $\mathbf{U}(t) = \mathbf{P}\tilde{\mathbf{U}}(t)$, whereas $\mathbf{P}$ is the squared non-singular matrix transformation of order N which should be calculated and $\mathbf{P}\tilde{U}$ is the vector of order N, known as the *vector of generalized displacement*.

The equation of motion in the new base is given by:

$$\tilde{\mathbf{M}}\ddot{\tilde{\mathbf{U}}} + \tilde{\mathbf{C}}\dot{\tilde{\mathbf{U}}} + \tilde{\mathbf{K}}\tilde{\mathbf{U}} = \tilde{\mathbf{F}}(t) \qquad (14.3)$$

In practice, an effective transformation is obtained by using the solutions to the equation of balance, assuming displacements of free vibration with $\mathbf{F}(t) = 0$. Obviously, if the system is balanced, $\mathbf{C}$ should be 0 so there is no loss of energy, thus resulting in the equation:

$$\mathbf{M\ddot{U}} + \mathbf{KU} = 0 \tag{14.4}$$

The solutions to this equation are harmonics of the type $\mathbf{U} = \phi \sin\omega(t - t_0)$, where ϕ is a vector of order N which indicates the extent of the vibration, t is the temporary variable t_0 is a time constant and ω is the frequency of the vibration of vector ϕ.

The solution is substituted in harmonic form in the equation put forth, the result will be a generalized eigenproblem from which we can determine ϕ and ω:

$$K\phi = \omega^2 \mathbf{M}\phi \tag{14.5}$$

The solution to this eigenproblem (Press et al. 1992) is made up of $3n$ eigensolutions.

If we define a matrix Φ whose columns are eigenvectors ϕ_i and a diagonal matrix Ω^2 which has the ω_i^2 eigenvalues on its diagonal, then the $3n$ solutions can be written as follows:

$$K\Phi = M\Phi\Omega^2 \tag{14.6}$$

Matrix Φ will be an appropriate transformation matrix (known as P), and therefore the expression is as follows:

$$\mathbf{U}(t) = \Phi\tilde{\mathbf{U}} = \sum_{i=1}^{N} \tilde{u}_i(t)\phi_i \tag{14.7}$$

where ϕ_i is the i–mode, $\tilde{u}_i$ is the amplitude and ω_i is its frequency.

In fact, an approach to the nodal displacements $\mathbf{U}(t)$ as $\hat{\mathbf{U}}(t)$ is done, the sum of p modes with lower frequency, being $p \ll N$.

$$\hat{\mathbf{U}}(t) = \sum_{i=1}^{p} \tilde{u}_i(t)\phi_i \tag{14.8}$$

Vectors $(\phi_i)_{i=1,\ldots,p}$ form a *reduced modal base of the system* and allow a closed form solution through the selection of modes with lower frequency.

Analytical Calculation of the Proper Vibration Modes

Once the elastic structure has been defined, it is necessary to present an analytical expression which allows its calculation to be carried out. Although it is possible to

directly use the numeric procedures to do so (14.6), (Nastar 1994) proposes a calculation to determine the analytical expression of the modes and, thus, considerably reduce its computational cost.

In the case of a closed curve, the eigenvalues ω^2 can be defined by the dispersion equation:

$$\omega^2 = \frac{4\,K}{M} \sin^2\left(\frac{p\pi}{n}\right) \tag{14.9}$$

For the simple eigenvalues, the expression of the associated eigenvector is given by:

$$\phi = \left[\cos\frac{2p\pi}{N}, \ldots, \cos\frac{2p\pi n}{N}, \ldots, \cos\frac{2p\pi N}{N}\right]^T \tag{14.10}$$

where $p \in \mathcal{B}(N)$, the *first Brillouin zone* (Nastar 1994).

For double eigenvalues, a couple of orthogonal eigenvectors is obtained, given by:

$$\begin{cases} \phi = \left[\cos\frac{2p\pi}{N}, \ldots, \cos\frac{2p\pi n}{N}, \ldots, \cos\frac{2p\pi N}{N}\right]^T \\ \\ \phi' = \left[\sin\frac{2p\pi}{N}, \ldots, \sin\frac{2p\pi n}{N}, \ldots, \sin\frac{2p\pi N}{N}\right]^T \end{cases} \tag{14.11}$$

Deformation Spectrum and Its Application

Once it's defined an analytical expression of the vibration modes of the nodes located on a curve, it defines a representation mechanism of the modal width value as a function of the shown mode: $\tilde{u}_i(t) = f(i)$

This representation will be called modal spectrum (Nastar 1994). Basically, it describes which modes and in which quantity are excited in order to alter a shape until it matches another. Thus it obtains an indictor of the quantity of *deformation energy* needed for the process:

$$E_{\text{deformation}} = \frac{1}{2}\mathbf{U}^T\mathbf{K}\mathbf{U} = \frac{1}{2}\sum_{i=1}^{N}\omega_i^2\tilde{u}_i^2 \tag{14.12}$$

Two deformations will be *similar* when their corresponding displacement fields $\mathbf{U}_1$ and $\mathbf{U}_2$ and similar, except the rigid transformations, which contain zero deformation energy. In other words: *two deformations will be similar when their corresponding deformation spectra are similar.*

Similarity Measure Between Complete Vessels

For the comparison of the complete profiles, being that we already know the number of points of the simple, a reference prototype is used. Each profile can be classified in relation to it's *deformation spectrum* to the prototype. This way, similarity between two profiles can be computed as the Euclidean distance between their associated deformation spectra.

$$d(D_1, D_2) = \frac{1}{p} \sqrt{\sum_{i=1}^{p} (\tilde{u}_i(D_1) - \tilde{u}_i(D_2))^2} \qquad (14.13)$$

Profile Scaling

The prototype C (Fig. 14.4) is the circumference centred in (0.5, 0.5) and that passes through the point (0, 0) subsampled uniformly in N points. All the profiles should be scaled in relationship to the prototype. This makes our *measure of deformation* invariant to scale.

First the profile P is aligned over the profile to calculate the spectrum (Fig. 14.5). The lowest point of the axis of rotation is aligned with the point (0, 0). Next, P is scaled uniformly so that its highest point corresponds with the edge of the piece that belongs to the circumference C.

Fig. 14.4 Circumference used as prototype
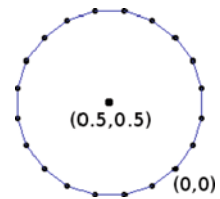
Fig. 14.5 Subsampled profiles and prototype
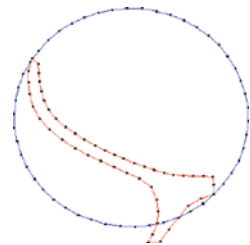
Contour Sampling and Node Correspondence

Different techniques are designed to establish the correspondence between the points (nodes) of the profile and the points of the prototype. The best results have been achieved when the profile is divided in exterior and interior halves that connect the origin with the edge of the piece. Both curves have been subsampled uniformly in *N/2* points (Fig. 14.6).

Fig. 14.6 Correspondence between profile and prototype nodes
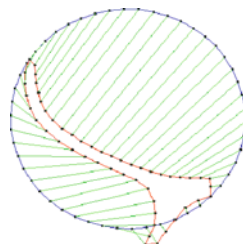
Similarity Between Fragments and Complete Profiles

The comparison of fragments with complete profiles can not be done with a prototype. This is because we do not know the number of nodes of the piece that correspond to the fragment points. Thus a different method must be used.

In the first place, three significant points are designated on any profile, called *anchor points* (Fig. 14.7). These anchor points are the edge, the lowest point that is in contact with the axis, and the base point most distant with the axis.

To be able to compare a fragment, it's necessary that it contains at least one of the three anchor points, as well as knowing the scale ratio between the associated fragment image and the complete profile. The whole process is as follows:

- Adjust the profile to a known or standard scale. The fragment should be scaled proportionally in relation to the scale of the profile.
- Measure the length of the contours of the fragment, to each side of each anchor point contained in the fragment.
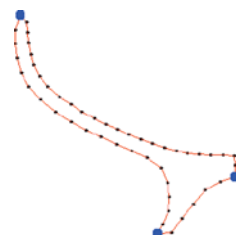
Fig. 14.7 Anchor points

- Take out the portion of the profile contours that correspond to the same lengths measured in the fragment, in relation to the anchor points.
- Subsample both contours in N points, with each side of each anchor point uniformly distributed and so that each point that corresponds with and anchor point is in the same relative position.
- Calculate the spectrum of deformation for open chains.
- Calculate the similarity between them (Eq. 14.13).

Experimental Results

We has been tested the output of the comparison scheme against a given set of 121 expert made, complete profile classification of examples taken from the database. As the information about fragments hasn't been allowed, the results shown here only refer to complete profiles.

Each expert, given a test profile T, has marked as *relevant* a variable size subset of the database profiles. Next, we have computed the similarity measure, using 200 nodes per contour, for T and the rest of the profiles in the database, and ordered increasingly the results. We has been computed two different classification performance estimators, applied to the first N profiles returned by our algorithm, given T:

- *Precision*:

$$P = 100 \cdot \frac{R_N}{N}$$

- *Recall*:

$$R = 100 \cdot \frac{R_N}{N_T}$$

being N_T the total number of relevant profiles related to T, and R_N the number of relevant profiles present in the first N returned profiles. Both measures range from 0 (worse) to 100 (best). For each example profile, it's computed the maximum obtained precision P_{max}, the N value that gives such precision (N_{Pmax}), and the minimum value of N that gives $R = 100$, (N_{Rmax}). The average values obtained are: P_{max} = 46.67%, N_{Pmax} = 25, and N_{Rmax} = 121.

Table 14.2 represents the accumulated percentile values for P_{max}, N_{Pmax}, and N_{Rmax}. For example, the column labeled 50% means that for one half of the profiles present in the database, a query with $N = 7$ returns at least a 45.5% of *relevant* profiles, and a query with $N = 28$ provides all of the *relevant* ones. Figures 14.8 and 14.9 show some classification examples. It must be noted that the classification performed by the experts does not only attend to geometrical cues, but also to semantic ones, therefore the algorithm can return profiles that are geometrically similar to the query one, but unrelated from the expert point of view.

Table 14.2 Accumulated percentile values for P_{max}, N_{Pmax}, and N_{Rmax}

Percentile	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
P_{max}	100.0%	75.0%	60.0%	50.0%	45.5%	33.3%	25.0%	16.7%	4.8%	0.2%
N_{Pmax}	1	3	4	5	7	9	12	17	29	568
N_{Rmax}	1	4	9	14	28	44	126	233	410	874
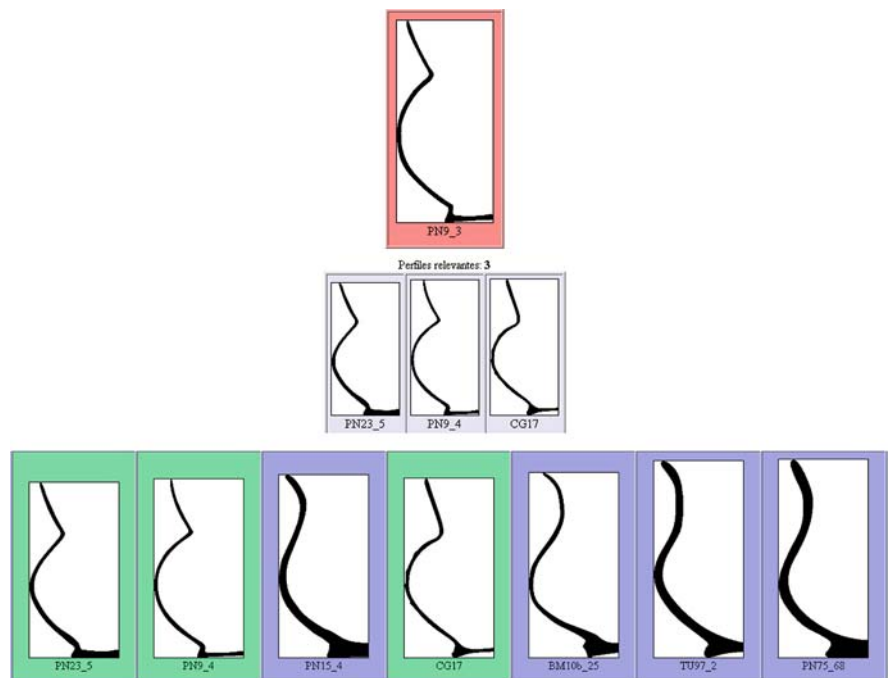
Fig. 14.8 First row: query image. Second row: relevant profiles. Third row: classifications results (relevant profiles marked in *light grey*)
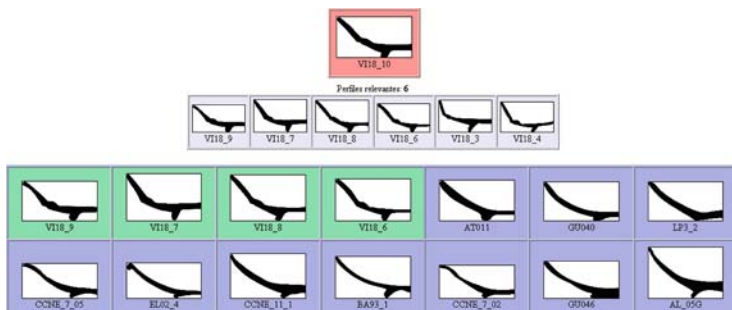
Fig. 14.9 First row: query image. Second row: relevant profiles. Third and fourth row: classifications results (relevant profiles marked in *light grey*)

Conclusions

Archaeology is a comparative discipline: archaeologist spends much of their time making comparisons between artefacts, between assemblages, between sites and even between regions. Such comparisons may be made in different ways for different purposes; for example, assemblages may be compared for chronological purposes (e.g. seriation), for spatial purposes (e.g. distributional studies), or for the study of function or status. None of these activities is possible unless comparable "objects" (in the widest sense) are given the same names wherever they occur. If different terminologies are used in different places, then a means of translating between them must be made available (Orton 2005). This generates anarchic situations like the different nomenclature used for the same shape of a vessel.

The methodology exposed constitutes a multidisciplinary approach for the anarchic situation concerning to the elaboration of a ceramic typology due to:

- The homogenization of the drawing.
- Uniformity of the methodology used.

Also, with the results exposed it is possible to say that in the majority of the cases the classification made for an archaeologist and the results obtained with this methodology coincide.

References

Adams, W.Y. Archaeological classification: theory versus practice. *Antiquity*, 62(234):40–66, 1988.

Bathe, K.J. *Finite element procedures in engineering analysis*. Prentice-Hall, Englewood Cliffs, 1982.

Braun, D.P. *Pots as tools*. In *Archaeological hammers and theories*. J.A. Moore, A.S. Keene (Ed.), Academy Press, New York, 1983.

Butzer, K.W. *Arqueología: una ecología del hombre: método y teoría para un enfoque contextual*. Barcelona, Bellaterra, D.L., 2007.

Clarke, D.L. *Arqueología analítica*. Ed. Crítica. 1984.

Lustig-Arecco, V. *Technology: Strategies for survival*. Holt, Rinehart and Winston, New York, 1975.

Melero, F.J., Torres J.C., León, A. On the interactive 3D reconstruction of iberian vessels, In *4th International Symposium on Virtual Reality, Archaeology and Cultural Heritage, VAST03*. F. Niccolucci; D. Arnold, A. Chalmers (Eds.), 2003.

Nastar, C. *Modèles physiques déformables et modes vibratoires pour l'analyse du mouvement non-rigide dans les images multidimensionelles*. PhD thesis, INRIA, 1994.

Nastar, C., Ayache, N. Frequency-based nonrigid motion analysis: Application to four dimensional medical images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(11):1067–1079, November 1996.

Pereira, J. La cerámica ibérica procedente de Toya (Peal de Becerro, Jaén) en el Museo Arqueológico Nacional, *Trabajos de Prehistoria, 36*, Madrid, 189–348, 1979.

Orton, C., Tyers, P., Vinci, A. *Pottery in archaeology*. Cambridge University Press, Cambridge, UK, 1993.

Oswalt, W.H. *Habitat and technology*. New York: Holt, Rinehart and Winston. 1973.

Pellicer, M. Las cerámicas del mundo fenicio en el bajo Guadalquivir. Evolución y cronología según el cerro Macareno (Sevilla), *Madriger Beitragë*, 8, 1982.

Pentland, A., Sclaroff, S. Closed-form solutions for physically-based shape modeling and recognition. *IEEE Trans. on Patterns Analysis and Machine Intelligence*, 13(7):715–729, July 1991.

Orton, C. Some thoughts on the history of reference collections in the UK. In Lange, A.G. (ed.). *Reference Collections Foundations for Future Archaeology*. Amersfoort: ROB, 23–34, 2005.

Pereira, J. La Cerámica ibérica de la Cuenca del Guadalquivir. I, Propuesta de clasificación. *Trabajos de prehistoria*, 45. 1989.

Pla, E., Aranegui, C. *La cerámica ibérica, en la baja época de la cultura ibérica*. Madrid. 1981.

Press,W.H., Teukolsy, S.A., Vetterling, W.T., Flannery, B.P. *Numerical recipes in C. The art of scientific computing, second edition*. Cambridge University Press, Cambridge, 1992.

Ruiz, A., Molinos, M. *Los iberos: análisis arqueológico de un proceso histórico*. Editorial Crítica, Barcelona, 1993.

Ruiz Mata, D. La formación de la cultura turdetana en la bahía de Cádiz a través del Castillo de Doña Blanca. In *I Jornadas de Arqueología del mundo ibérico (1985)*, Jaén, 1987.

Ruiz Rodriguez, A., Molinos, M. et al. El horizonte Ibérico Antiguo del Cerro de La Coronilla. (Cazalilla, Jaén). Cortes A y F. *Cuadernos de Prehistoria de la Universidad de Granada*. 8:251–295, 1983.

Shennan, S. *Arqueología cuantitativa; traducción castellana de Juan Antonio Barceló*. Editorial Crítica, Barcelona, 1992.

Shepard, A. *Ceramics for the archaeologist*. Carnegie Institution of Washington, Washington, DC, 1956.

Chapter 15
Some Applications of Geometric Morphometrics to Archaeology

Marcelo Cardillo

Idea and Aims

This work explores some aspects of the application of geometric morphometric techniques in archeology, with a focus on lithic artifacts. We show that Elliptic Fourier Analysis and landmark/semi-landmark based methods can easily generate quantitative useful information relative to outline variation in lithic artifacts. This information can be used latter as raw data into univariate, multivariate analysis to explore mayor trends of morphological variation as well as relations between metric and morphological variation.

Introduction

As in other disciplines that used classification procedures, archeology depends heavily on classification to analyze and explain variation. However, as Gero and Mazullo pointed out (1984), many traditional typologies are based on an intuitive recognition of patterns, where types are defined as a series of idealized forms, broken down into subvarieties on the basis of some number of defining variables.

This selection criterion is often biased, and the analysis cannot be replicated by other researchers. Dunnell (1971) observed that common typological analysis based on invariable properties of artifacts, make difficult the study of change, and referred these to an essentialist typology, contrary to materialist one (see also Hiscock 2001). A materialist approach to variation emphasizes a statistical treatment and management of data. Classification and analysis in lithic technology is commonly based on discrete, qualitative traits. Often, the classes or types are generated cutting down continuous metric and morphological variation into varieties or subclasses. These divisions are at last, arbitrary actions, which increase intra-observer error among

M. Cardillo (✉)
IMHICIHU-CONICET, Saavedra 5th floor, Buenos Aires, Argentina
e-mail: marcelo.cardillo@gmail.com

lithic analysts and impede or difficult the replication by other researchers (see extended discussion in Dunnell 1971, Hiscock 2001).

We believe that the geometric morphometric approach is related with a more materialistic view of technology where the focus is on continuous quantitative phenomena rather than qualitative. Also, a more integral approximation to artifactual variation result from using geometric morphometrics tools for description, classification, and analysis. In this sense, artifactual variation can't be seen as self-explained phenomena but linked to different factors in need to be explained in each case. According to Shott (1996) variation within a single class or artifacts may be related to random sources like individual variation, style, replicative error, raw material variation and measurement error. Also, other factors, like manufacture and performance criteria are related to function (sensu Dunnell 1978).

In archaeology, a traditional morphometrics (*sensu* Marcus 1990) approach to lithic analysis was implemented through linear measurements as length, width, thickness, ratios and angles (see Wynn and Tierson 1990, Crompton and Gowlett 1993, Franco et al. 2005 among others), also three dimensional scanning techniques were used (Grosman et al. 2008). A primary concern of this analysis was to measure within class variation or morphological changes due to use and reactivation of artifact edges (Hiscock 2003; Hiscock and Clarkson 2005; Buchanan 2006; Shott et al. 2007; among others). In recent years, morphometric techniques based on different geometric models became more common, although in very different ways. These disparate approximations prevent the development of a common language for shape studies in archaeology, and also discourage researchers who want to start using these techniques.

In one of the first systematic applications of geometric morphometrics in archaeology, Gero and Mazzullo (1984) used elliptical Fourier analysis (EFA) over closed forms of lithic flakes for different time periods in Peru. These authors found that different levels of variation in harmonics amplitude describe changes in flake detachment techniques and relative standardization, observed as a paulatine angularity reduction trough time. In a similar fashion, Saragusti et al. (2005) shows the potential application of Fourier descriptors to make account of shape variation related to deformation, symmetry, roughness, and surface of different artifacts. In another work Saragusti et al. (1998) applied mathematical equations to study changes in symmetry, showing a temporal trend to more symmetric artifacts in lower Paleolithic handaxes. In relation to the use of landmarks, Brande and Saragusti (1996) defined important methodological issues related to the application of a landmarks based method to the study of artifacts. This works develops a geometric model to study handaxes, focused on linear measurements taken at regular intervals and then transformed into shape coordinates. In a similar fashion, Lycett et al. (2006) explores three dimensional morphometrics of Pleistocene lithic cores. The author takes several measures with a special purpose caliper and transformed them into shape coordinates, and after that submitted it to multivariate morphometric analysis. The results reflect the mayor trends of variation in lithic nuclei, as general dilation compression and relative asymmetry (also see Brande and Saragusti 1999 for an early exploration with three dimensional landmarks).

One of the most paradigmatic artifacts in lithic analysis is the projectile point. This kind of artifact was used to explore change in subsistence practices, stylistic or functional change among other approaches. Shape change in projectile points were accounted for Cardillo (2006) and Castiñeira et al. (2009) and Franco et al. (2009) among others. The existence of variations in the design of stemmed bifacial projectile points using geometric morphometric analysis, combined with linear measurements and microwear analysis was assess Franco et al. (2009). Results suggest that shape variation in the stem section of projectile points are not related to hafting technique defined by microwear analysis or metrical variation. Also, morphological change referred to resharpening and reavivation of artifact edges was explored within scrapers with EFA technique (Cardillo and Charlin 2009) and semi-landmarks (Cardillo 2009). In both cases, we found that variation display as a continuum is best explained by resharpening intensity and raw material acquisition and exploitation. Also, this variation can't be explained with a common typological approach. A common element in these studies is a focus on capturing variation of contours at various levels, using different parameters. It is important to note that lithic artifacts have common smooth contours of curves and plane convex or plane concave sections in essentially a two-dimensional outline. For this reason, outline description was a primary focus of inquiry in these investigations.

Of this different methods, we believe that landmark and semi-landmak approximations (Bookstein 1991, 1997) or EFA (Kuhl and Giardina 1982; Rohlf 1990), have more potential in the study of artifactual variation because they are based on easy to learn steps and comprehensive free software as Morpheus et al. (Slice 1998), Tps series (Rohlf 2002a), IMP series (Sheets 2003) Past program (Hammer et al. 2001), among others. Landmark semi-landmarks and EFA based approach are very flexible tools to describe and visualize shape change in an interactive manner as continuum phenomena, based on a sets of digitized x/y coordinates or x/y/z in a three dimensional case. These methods may prove useful to study, among others, change related with artifactual edge rejuvenation (reactivation or resharpening) and morphological variation due to functional or performance requirements. Here we show some of the potential of these methods with simple examples where geometric morphometrics are used as a tool to study lithic technology in a more quantitative and detailed manner. In this sense, we believe that a major potential in morphometric application in archaeology, is linked to visualization and numerical description of outlines and therefore, to the use of semi-landmarks, and the EFA method. This is because in the lithic analysis, only few landmarks (or homologous points) can be defined in the sense of Bookstein (1991).

In archaeology, the landmarks are according to "type two" landmarks of Bookstein (1991), which define them as points located in the maximum of curvature or extreme points in morphology (Bookstein 1991; Zelditch et al. 2004). Neverless in some cases, these can be difficult to establish, because artifacts can show variable morphological attributes due to random or functional causes, as mentioned above. But the location of landmarks when is possible, can be very useful to take account shape change due to reactivation of artifact edges or projectile point tips as observed by Castiñeira et al. (2009). On the other hand, semi-landmaks are used to incorporate

information about outlines (Bookstein 1996/1997) defined as a set of points located at equal intervals along the curve. These points defined in terms of his relative position to other features (Zelditch 2004) in these cases, the entire outline are treated as a homologous unit. Therefore, relative variation to discrete semi-landmaks has no meaning per-se, and make sense when is studied as a whole. In relation of theoretical implications of the use of landmarks and outline descriptors, we consider that morphological features can be studied independently of homological information as common is used in biology (as Ferson et al. 1985 suggest, see also Bookstein 1996/1997). In fact artifacts not have biological homologies, but are the byproduct of recursive and standardized human technological practices, transmitted and maintained by cultural transmission and imitation (Cavalli-Sforza and Feldman 1981; Boyd and Richerdson 1995). However selection of outlines or discrete points must be related not only to research questions, but also to the particularities of each artifact topological feature as well as technical and morphological criteria, see also Brande and Saragusti (1996) and Lycett et al. (2006).

In this paper we focus on flaked lithic technology. Flaking artifacts from a piece of stone is a reductive process, where artifacts are made by removing flakes from a piece of parent lithic material. Different techniques as direct percussion flaking (striking the piece with a hammer) or pressure flaking (pressing a pointed instrument against the edge) are commonly used together to make different tools, from scrapers with steep-edge to knife with thinner edges and projectile points. Projectile points are only a part of more complex artifacts as throwing spears or bow and arrow technology. Also, lithic resharpening or reactivation practices that extend the use wear of lithic artifacts are reductive in nature. For this reason, reduction in size is a common byproduct that results in an allometric relation between form and size. Knapping processes themselves are subjected to random error related to rock texture, composition and grain, and also, knapper skills (Eerkens 2000; Eerkens and Bettinger 2001). For this reason, a different range of variation is expected even within the same kind of artifacts, variation being probably higher than the expectations, for example, in living organisms. Therefore, different types of artifacts will have a different rank of variation depending on the complexity of design, functional requirements or production techniques, and in the case of lithic artifacts the physical properties of the materials employed for knapping. For this reason, the potential discrimination between classes or subclasses of artifacts, or the power of multivariate analysis to explain the major trends of variation depends in some extent of the kind of artifacts analyzed and the selected methods to capture the morphological information. It is likely, that different kind of artifacts require different approaches, depending on their morphological features. In this regard, we believe that Fourier and landmarks and semi-landmarks based methods can give an efficient account of the shape variation in almost all cases.

To explore some applications of these methods in common lithic analysis and classification, we show three examples previously studied by Cardillo (2006), Scartascini and Cardillo (2009), and Castiñeira et al. (2009). The first two cases, use landmarks and semi-landmarks methods, and the third case, EFA approach.

Materials and Methods

First Case: Morphotype Variation in Simplest Outline: Line or Fish Weights

The line weight or net weight stones are artifacts commonly found in some areas in coastal north Patagonia (Sacartascini and Cardillo 2009) and are related to the exploitation of marine resources. Little archaeological information of fishing techniques exists, due to the fact that only weights were preserved. These artifacts were made with little modification of the original piece of stone, using pebbles from gravel deposits located near the sea shore (Scartascini and Cardillo 2009). Artifact manufacture was limited to knapping two notches in each extreme of the pebbles.

The sample is composed by 56 artifacts from three archaeological areas located along the north coast of San Matías Gulf, río Negro, República Argentina (Fig. 15.1a). Given the little energy investment in these artifacts, our primary interest was to obtain exhaustive characterization morphological variation relative pebble selection criteria. To that end, we measured metric variables as length, width, thickness and weigh, as well as the size of the notches in each of the ends, in order to explore correlations between shape and size.

Fig. 15.1 Geographical location of samples (**a**) north coast of San matías Gulf, río Negro, República Argentina, (**b**) República Oriental del Uruguay y (**c**) Puna of Salta, República Argentina

Second Case: Allometric Change in Paleoindian Projectile Points "Fishtail" from Uruguay

24 instruments classified as paleoindian Fishtail projectile points (around 11–10 years Ka B.P.) from surface collections of different localities in the República Oriental del Uruguay were analyzed. They are stored in public and private collections (Fig. 15.1b). Available radiocarbonic chronology (Nami 2007) supports the statement that the "Fishtail projectile point" morphotype is related to first human occupations processes in South America (during Pleistocene–Holocene period). In this case, we use geometric morphometric to make account to the allometric process of shape change related to blade rejuvenation of projectile points using centroid size of digitized images as a measure of size change.

Third Case: Projetile Point Change in Archaic Period in Salta. Puna Region

The sample was obtained from surface and sub-surface contexts in the Ramadas site, located in San Antonio de los Cobres valley, Puna of Argentina (Fig. 15.1c). The temporal span is between 6000 B.P. and 4000 B.P. The technological sequence of this site is similar to others recorded in the dry and salty Puna, for example in sites as those from Quebrada Seca and Inca Cueva 4 (Martínez 1999)

We selected a sample of ten morphotypes (or morphological variants); nine of them correspond to the samples collected in the studied zone, but one of these (the selected outgroup -OG- that dates from 6000 B.P.) proceeds from the site Quebrada Seca 3 excavated by Carlos Aschero (Aschero 1988; Aschero et al. 1993, 1994). The selected morphotypes are well represented in the archaeological sequences of very different sites with also good chronological information (see Ratto 2003; Martínez 1997). These authors suggest that metric variation was related with the use of different hunting weapons (spear-thrower weapon or more weighty hand thrower spears).

The main focus was to explore the temporal trends of change within this artifactual class, as well as the relation between metric and shape change.

In all cases, images were taken with an eight megapixels digital camera and no more than 30 cm of focal distance. Given that the analysis focused mainly on the contours of the artifacts, the images were taken in grayscale on a contrasting base to increase edge resolution. Before that, we used a variable number of sampling points around the outline using tpsDig (Rohlf 2002a) program using the automatic outline detection mode. All artifacts were recorded in a standard orientation, previously defined according to morphological and technological criteria (Fig. 15.2).

In the first case, where the artifacts show a considerable variation we use 100 closely spaced points, and the outline tracing began at the most distal point (Fig. 15.2a). In the second case, two landmarks and 22 semilandmarks was used. Landmarks were located at the tip and the base of specimens (Fig. 15.2b). Using the program Make Fan (Sheets 2003) equally angle spaced point were located using only a half portion of each artifact, in order to avoid information related to the

Fig. 15.2 The three examples presented in this work. (**a**) stone weight, (**b**) Fishtail projectile points (**c**) lanceolate projectile points. In (**a**) and (**c**) the *open circles* show the point were automatic digitalization begins in (**b**) the *open circles* indicate the location of the landmarks used, the shading area indicates the half portion selected to put semilandmarks
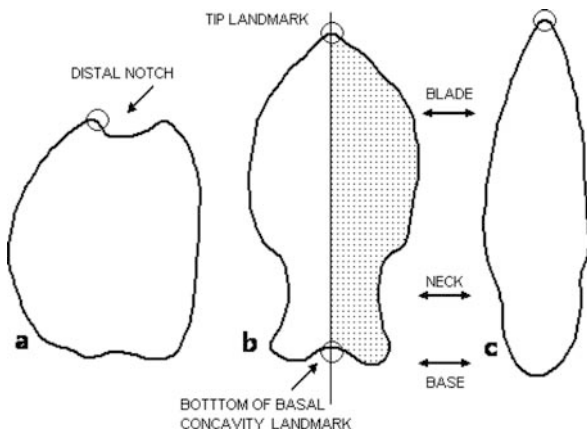
asymmetry. In this case we use Partial Least Square method (Rohlf and Corti 2000) to explore covariation between set of metrical attributes and shape using TpsPLS (Rohlf 2002a) as maximum length, thickness, width, and centroid size of specimens. Centroid size is the square root of the summed squared distances of each landmark form the centroid of the landmark configuration, and is obtained from the images in the process of superimposition. Correlation of shape change and centroid size is a good method to explore allometric change, as expected in the case of reactivation as in the case that Shott et al. (2008) shows.

Once a set of points around the outlines was digitized, landmarks and semilandmarks were processed using a generalized Procrustes analysis (Rohlf 1999), Also, in order to reduce de effect relative to the they arbitrary position, semilandmarks were aligned using bending energy minimization criteria (Bookstein 1997). To explore mayor trends in morphological variation, the resulting shape coordinates were submitted to relative warp analysis using TpsRW program (Rohlf 20002a) that are principal components of the partial warp shape variation at different scales (Rohlf 1993). An important aspect of relative warps is that the results of statistical analysis can be expressed as an intuitive deformation grid diagram of each case with respect to the mean form or reference.

In the third case, we use 100 equally spaced points along one smooth curve (Fig. 15.1). Digitalization was also made with automatic outline detection utility in tpsDig program. In this case, digitalization begins at the tip of projectile. Resulting coordinates were submitted to EFA using the program Past (Hammer et al. 2001). EFA method fitting successive sine and cosine terms (harmonics) these harmonics decreasing in amplitude to the first (lower) to higher harmonics. These harmonics describe components of shape at different scales (Rohlf 1990). In this case the first 20 harmonics were then using in principal component analysis to reduce dimensionality. Also the first principal component axes that explain mayor trends of morphological variation in outline are used as new variables in regression analysis.

Also, to explore grouping patterns and historical relations between projectile points we use the neighbor joining method (NJ). This method was proposed by Saitou and Nei to analyze distance data (Saitou and Nei 1987). This procedure,

generate phenetic trees from continuous data like morphological multivariate data. NJ tree is an unrooted tree, but can be rooted like parsimony based methods. The input data to NJ procedure was the Fourier coefficients or harmonics obtained for each case. The earlier morphotype was use as outgroup to polarize the resulting phylogram. The likehood of resulting tree was computed by bootstrap (1,000 times). NJ method and bootstrap were performed using the program Past (Hammer et al. 2001). Also, general statistics, as principal component analysis, mixture analysis and regression were performed with the same software. Mixture analysis is an advanced maximum-likelihood method for estimating the parameters (mean, standard deviation and proportion) of two or more univariate normal distributions, based on a pooled univariate sample (Hammer et al. 2001).

Analyses

First Case

The RW analysis using semilandmarks show that the first component explains about 81% of shape variability, while the second 7, 48% (Fig. 15.3).

Given the focus on global description of morphological variation, the analysis proceeds with the uniform component, this describes the overall trends in compression/dilation or stretch of shape (Zelditch et al. 2004). For this reason the first RW shows that the greater variation is explained by big scale compression dilation patterns while the second RW shows that variation on asymmetry of pebbles used as weights. Also correlation was carried out between the first three RW and length, width, thickness, weight, but not significant correlations was observed in any cases.

The clustering distribution of cases inside the morphospace (concentration of the cases in two different clusters) of the first two RW, suggest a morphological gap or discontinuity. Finally, we use mixture analysis on the first RW in order to explore if that pattern can be best explained by the existence of two different distributions. The results show that the two group's hypotheses have the best likelihood score (Fig. 15.4).

Results suggest that a different selection pattern of pebbles was carried out by humans, although no relation between shape and size variation was observed. It appears that, morphological variation responds first to performance requirements of these artifacts related to hydrodynamic requirements. Due to little modification, metric variation in natural outcrops of lithic deposits, may explain much of the morphological variation observed here.

Second Case

The RW analysis shows that the first component explains 56% of shape variation (Fig. 15.5). Variation was explore with and without the uniform component (Rohlf
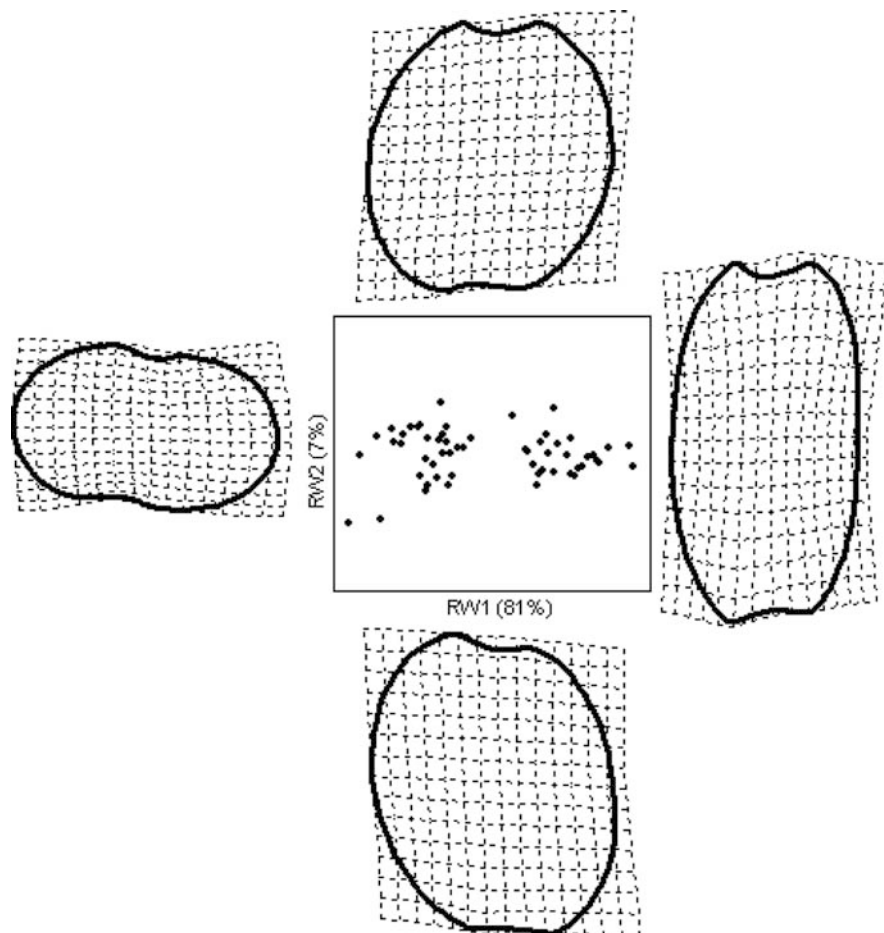
Fig. 15.3 First two axes of RW analysis of stone weights

and Bookstein 2003) and employing a variable weight to partial warps at different scales Rohlf (1993) alpha $= 0$ (gives equal weight at different spatial scales), and alpha $= 1$ (that gives more emphasis to variation at larger spatial scales). In all cases we got similar results.

The first axes of RW analysis shows the relative dilation/compression of blade and neck of projectile points. Also no discontinuities are observed in the first morphospace distribution; this pattern suggests a continuum of morphological change. To explore if this pattern was related reactivation/rejuvenation of blade a multiple regression using different variables was carried out, including the uniform component with partial least square method. Through this analysis is observed a significant correlation ($r = 0.65$, $p < 0.05$) between the centroid size and blade shape which indicates an allometric relationship between shape and size (Fig. 15.6).

Fig. 15.4 Mixture analysis plot showing two slightly overlapped distributions using the first RW
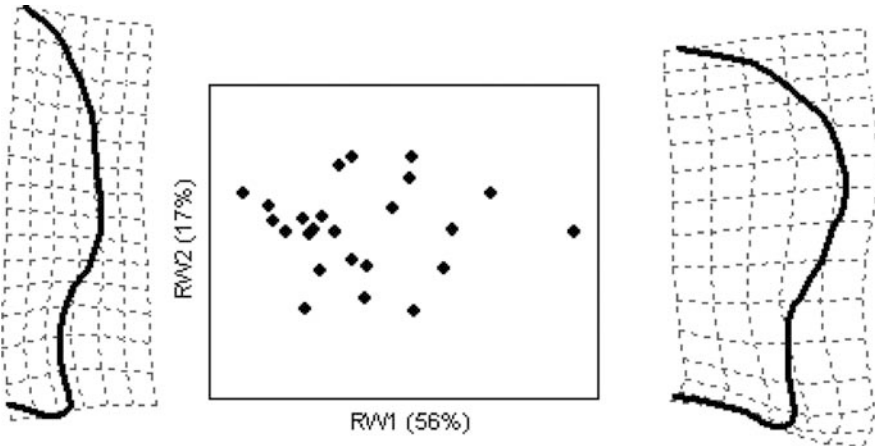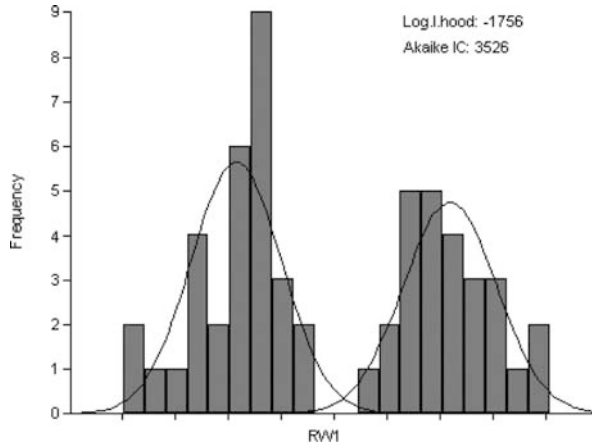
Fig. 15.5 Two first RW showing thin plate spline, including uniform component and alpha $= 0$

The analysis suggests that projectile points became more rounded while the geometric size decreases. At the right of Fig. 15.6 shows the shape relative to smaller artifact as deformation grids (a) and also display them as by vectors of relative landmark displacements (b). In both cases we can see the pattern and direction of allometric change in witch the blade is contracted in relation to the expansion of neck, also affected by reactivation. Results suggest that morphological change is related to rejuvenation of Fishtail projectile points, resulting in allometric patterns as Shottt et al. (2007) observed in Folsom Points.

r=0.65 p<0.05
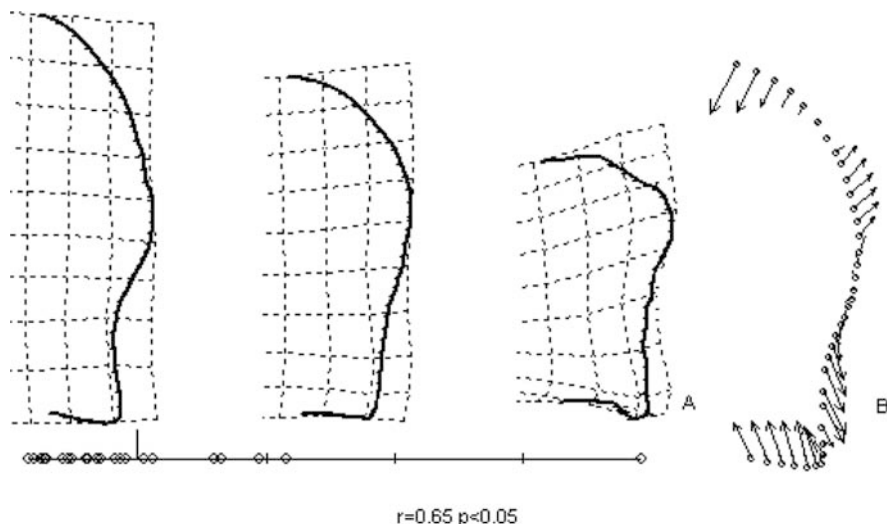
Fig. 15.6 Correlation between shape and centroid size. More exhausted or reactivated projectile points are toward the *right* of the figure variation (Fig. 15.6)

Third Case

The results of PCA over 20 harmonics shows that the first component explains an 88% of the total of the variance, while the second a 5.81%. The first PCA axis shows the general rounding or elongation of shape, while the variation of the remaining components is linked to relative asymmetry and shape change in more local scales. To explore how this change is related with an allometric pattern between size and shape, we made a regression analysis between weight (log) and the first PCA. Regression shows a positive correlation of $r = 0.66$ $p = 0.03$, in which more elliptical shapes are more light than elongated oval ones (Fig. 15.7).

To explore the pattern of morphological change from the oldest known morphotype, we use NJ method two perform a phylogram using the 20 harmonics as input. The results suggest a gradual trend of morphological change (Fig. 15.8), a same result can be seen in the first two axis of PC analysis (at the right, above)

Discussion

In the first case, results suggest that semi-landmarks based techniques are useful to capture the main trends of morphological variation, even in cases of highly variable shapes. We can also see potential discontinuities in the morphospace that may be related to statistical subclasses in Dunnell's (1971) sense. These subclasses can be explored with different statistical methods as mixture analysis or clustering

Fig. 15.7 Regression
between the weigh and the
first PC axis of EFA
coefficients
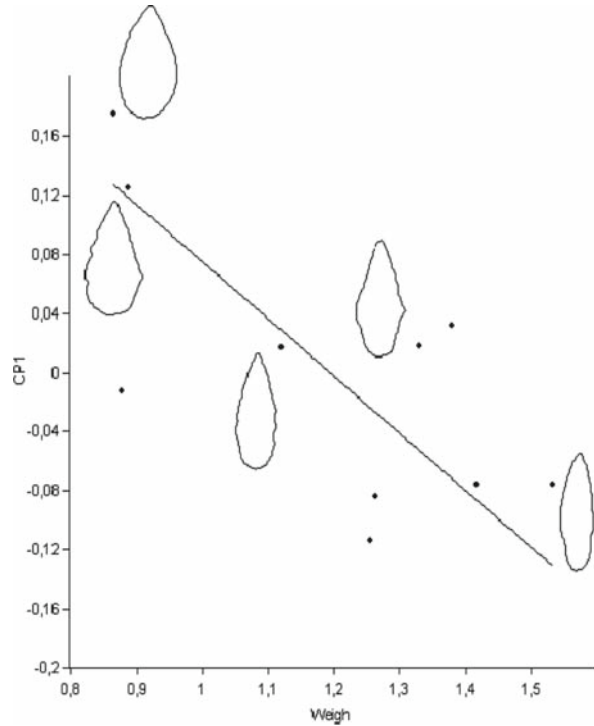
algorithms, as K-means. Also, we observed that particular morphotypes of rel-
ative symmetric pebbles was preferred. This selection criterion can be explored
for example, dividing each case in the middle and then performing two separate
morphometric analysis and shape versus shape regression with partial least square.

In the second case, employ landmarks and semi-landmarks together, gen-
eral trends of change was captured, but no discontinuities in morphospace were
observed. That can be related as a continuum of shape change inside the same basic
design. Also, the observed allometric relations between centroid size and shape
change, (almost located in the blade and neck areas), suggest that morphological
change is related at last some extent, with reactivation processes (Castiñeira et al.
2009).

While not shown here, in the first and third cases we use the previously aligned
points to perform Principal Component Analysis on EFA coefficients and RW anal-
ysis. The resulting coordinates of EFA and RW ordinations for fish weights and
lancolate points were compared by means of procrustean superimposition using
PROTEST (Peres Neto and Jackson 2001) through 10.000 permutations (results,
m12 = 0.87 $p < 0.001$ in the first case and m12 = 0.88 $p < 0.001$ in the second one).
These results suggest that similar ordinations or clustering patterns between cases
could be obtained by means of both methods. While our these results are very crude,
we found that both EFA and landmark/semi-landmark based methods give similar
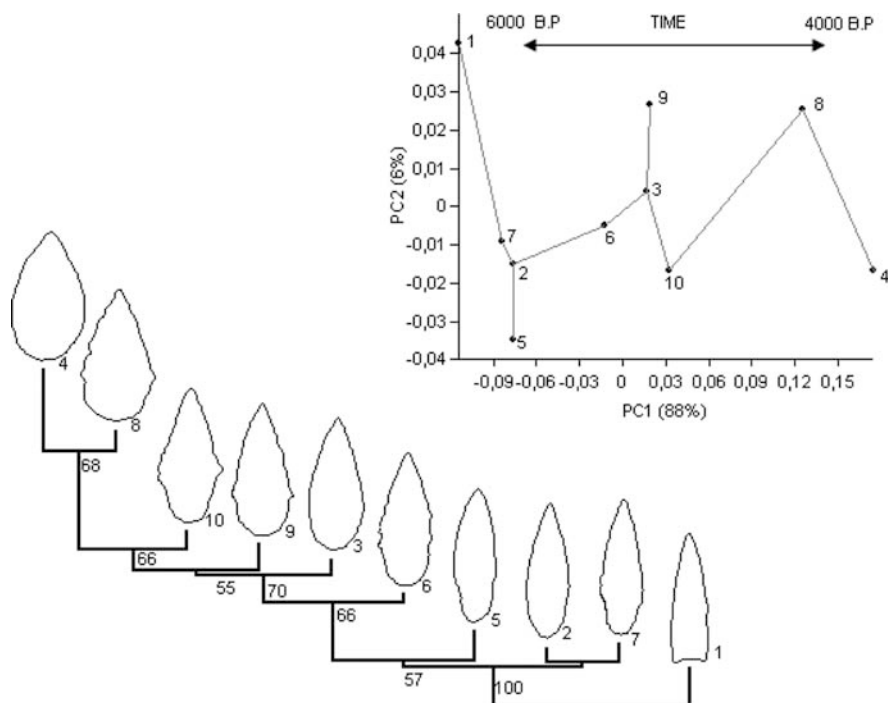
Fig. 15.8 Phylogram showing the clustering pattern of ten morphotypes. Numbers display de bootstrap support or each node only one case (2 and 7) was lower than 50%. At the *upper side* of figure we display scatter plot from PCA, cases were connected by a minimum spanning distance. *Arrows* shows the temporal trends

results in capture mayor trends of variations in two dimensional outlines taken as a whole, according to Sheets previous results in biological shapes (Sheets 2006).

Fourier harmonics (Rohlf 1998) or partial warp scores (Rohlf 2002b) can be used into clustering algorithms to explore morphological patterns (in witch some selected morphology or mean morphology can be used to rooting the tree). Also the morphospace generation and visualization with Thin Plate Spline or other methods can be use as a heuristic tool to explore variation patterns in different scales (Bookstein 1991). After that, different correlation/regression routines can be made, to pursuit the proximate causes of shape variation.

Because limited points can be used as landmarks in lithic artifacts, it appears that the common rule is a larger number of semilandmarks than landmarks. For this reason in almost all the lithic analysis the semi-landmarks have more weight in the results, as Sheets (2004) shows (see also Zelditch et al. 2004 for more complete discussion of this issue). One possibility is to divide morphology into a set of modules based on morphological of technological criteria. Correlation patterns between these modules (for example between the blade and de neck/base of a projectile point) can be related to functional integration of different sections of artifacts.

On the other hand, partitioning morphology into modules allows reducing the effect of sections with greater number of semilandmarks and therefore, more potential weigh in the results, as Sheets (2004) suggest. Also bounded regions of morphology and open outlines (as we show in the second case) can be more easily explored by landmark/semi-landmarks methods (see Franco et al. 2009) although some Fourier derived techniques can be used for open outlines as well.

Maybe one useful way to select between EFA or landmark methods depends on the nature of the data. The first method is better for complete outlines, and when scarce or no landmarks can be recognized or used. The landmarks/semilandmarks approach can be used for bounded regions of morphology, open or closed outlines, in this last case, with similar results.

Another important factor that we see is related with curation and reactivation of artifacts. This is a very common factor that can be expected to alter size and shape, result in allometric deformation, which is best explained as big scale shape deformations and uniform component related variation (compression/dilation and shear). Also we found that small scale variation along the outlines was related to roughness of lithic artifact as flaking or retouching of edges and microfractures due to taphonomic history (pot-depositional processes as trampling, abrassion, and weathering) of artifacts. These processes are taken into account when changes in roughness in one of the focus of analysis (see Gero 1984; Saragusti 2005) for this reason variation in local scale along the outline may be less informative than macro-scale variation. Small scale morphological change can be observed in some cases. But there is not a one single method for all possible cases, and much more work will be done with different kind or artifacts.

An other important factor that must be taken into account is that in archaeology, sample size are commonly small in relation to the number of variables as Fourier descriptors or semiladmarks; witch in turn can impede the use of some statistical methods, as canonical variation or discriminant analysis. One good possibility is using the first PCA axis of EFA series or first RW of partial warps (in this case, only the mayor portion of all variation selected). This axes can be used in univariate regression with independent variables (as weight of specimen) or in common correlation routines, as we show in the examples.

Finally, geometric morphometrics has many applications that go beyond shape analyses of lithic technology, different kind of archaeological data can be studied, and other variables can be used as well. Also, it would be useful to increase de interaction between researchers working on morphology through special purpose workshops and congress. This would help to the development of a common language morphometrics in archeology.

Conclusions

We think that geometric morphometric is a fertile ground to archaeology and can be part of a common protocol lithic study. This method brings us to powerful tool to explore and analyze variation, also implies theoretical and methodological

approximation to more materialistic approach to variation. In the case of lithic analysis, this allows quantitative description of variation of shape and more objective and testable results. Also numerical treatment of data can be used to explore design and performance hypothesis as well as temporal and spatial patterns and live history of artifacts.

References

Aschero CA (1988). El sitio ICC4: un asentamiento precerámico en la Quebrado de Inca Cueva (Jujuy, Argentina). Estudios Atacameños 7

Aschero CA, L Manzi and AG Gomez (1993–1994). Producción lítica y uso del espacio en el nivel 2b4 de Quebrada seca 3. Relaciones 19. Sociedad Argentina de Antropología. Buenos Aires

Bookstein FL (1991) Morphometrics tools for landmarks data. Geometry and biology. Cambridge University Press. NY

Bookstein FL (1996/1997) Landmarks methods for form without landmarks: morphometrics of group differences in outline shape. Medical Image Analysis. Oxford University press. 1(3) 225:243

Brande S and I Saragusti (1996) A Morphometric model and landmark analysis of acheulian handaxes from norhern Israel. Advances in morphometrics. Edited by LF Marcus, M Conti, A Loy, GJP Naylor and DE Slice. Plenum, New York

Boyd R and PJ Richerson (1995) Culture and the evolutionary process. University of Chicago Press. Chicago

Brande S and I Saragusti (1999) Arti-Facts Graphic visualization of Handaxes and other atifacts. Near Eastern Archaeology 62(4) 242:245

Buchanan B (2006) An analysis of Folsom projectile point resharpening using quantitative comparisons of form and allometry. Journal of Archaeological Science 33:185–199

Cardillo M (2006) Explorando la variación en las morfologías líticas a partir de las técnicas de análisis de contornos. El caso de las puntas de proyectil del holocenomedio-tardío de la Puna de Salta (San Antonio de los Cobres, Argentina). Werken 7, 77:88

Cardillo M and J Charlin (2009) Tendencias observadas en la variabilidad de los raspadores de norte y sur de Patagonia. Explorando las interrelaciones entre forma, tamaño e historia de vida. Arqueometría Latinoamericana 2do. Congreso Argentino y 1ro Latinoamericano. Vol. 2, 351–359. Buenos Aires. Argentina

Cardillo M (2009) Variabilidad en la manufactura y diseño de artefactos en el área costera patagónica. Un enfoque integrador Doctoral dissertation. Facultad de Filosofía y Letras. Universidad de Buenos Aires. MS

Castiñeira C, M Cardillo, J Charlin, J C Fernicola and J Baeza (2009) Análisis morfométrico de cabezales líticos "cola de pescado" de la Rep Oriental del Uruguay. Arqueometría Latinoamericana. 2do. Congreso Argentino y 1ro Latinoamericano. Vol. 1, 360–366. Buenos Aires. Argentina

Cavalli-Sforza L and MW Feldman (1981) Cultural transmission and evolution: a quantitative approach. Princeton University Press. Princeton

Crompton RH and JA Gowlett (1993) Allometry and multidimensional form in Acheulian Bifaces from KilombeKenya. Journal of Human Evolution 25, 175:199

Dunnell, RC (1971) Systematics in prehistory. Free press, NY

Dunnell RC (1978) Style and function: a fundamental dichotomy. American Antiquity 43, 192:202

Eerkens JW (2000) Practice makes whithin 5% of perfect: the role of visual perception, motor skills, and human memory in artifact variation and standardization. Current Anthropology 41, 663:668

Eerkens JW and RL Bettinger (2001) Techniques for asseing standardization in artifact assemblages. Can we scale material variability? American Antiquity 66(3) 493:504

Ferson S, FJ Rohlf and R Koehn (1985) Measuring shape variation of two-dimensional outlines. Systematic Zoology 34, 59:68

Franco NV, M Cardillo and LA Borrero (2005) Una primera aproximación a la variabilidad presente en las puntas de proyectil denominadas "Bird IV". Werken 6, 81:95

Franco NV, A Castro, M Cardillo and J Charlin (2009) La importancia de las variables morfológicas, métricas y de microdesgaste para evaluar las diferencias en diseños de puntas de proyectil bifaciales pedunculadas: un ejemplo del sur de Patagonia continental. Magallania 37(1) 99:112

Gero J and J Mazzullo (1984) Analysis of artifact shape using Fourier series in closed form. Journal of Field Archaeology 11(3) 315:322

Grosman L, O Smikt and U Smilansky (2008) On the application of 3-D scanning technology for the documentation and typology of lithic artifacts. Journal of Archaeological Science 35, 3101:3110

Hammer Ø, DAT Harper and PD Ryan (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4 (1) 9. http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Hiscock P (2003) Early Australian implement variation: a reduction model. Journal of Achaeological Science 30, 239:249

Hiscock P (2007) Looking the other way. A materialist/technological approach to classifying tools and implements, cores and retouched flakes. Tools versus Cores? Alternative approaches to Stone Tool Analysis. Edited by S. McPherron. Cambridge Scholars Publishing: 198–222

Hiscock P and C Clarkson (2005) Experimental evaluation of Kuhn's Geometric Index of Reduction and the flat-flake problem. Journal of Archaeological Science 32:1015–1022

Kuhl FP and CR Giardina (1982) Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18, 236:258

Lycett SJ, N von Cramon-Taubadel and RA Foley (2006) A crossbeam co-ordinate caliper for the morphometric analysisof lithic nuclei: a description, test and empirical examples of application. Journal of Archaeological Science 33, 847:861

Marcus L F (1990) Traditional morphometrics. Proceedings of the Michigan morphometrics workshop, Edited by FJ Rohlf and FL Bookstein, Special Publication Number 2. The University of Michigan Museum of Zoology. Ann Arbor. Michigan. 77:122

Martínez JG (1997) Estrategias y Técnicas de caza. Análisis tipológico-tecnológico de proyectiles arqueológicos. Trabajo Final de la Carrera de Arqueología. Facultad de Ciencias Naturales e Instituto Miguel Lillo. Universidad Nacional del Tucumán. MS

Martínez JG (1999) Puntas de proyectil, diseño y materias primas. En los tres reinos: prácticas de recolección en el Cono Sur de América. Edited by Por CA Aschero, MA Korstanje and PM Vuoto. Magna Publicaciones. Universidad Nacional del Tucumán. Argentina. 61:69

Nami H (2007) Research in the middle Negro river basin (Uruguay) and the paleoindian occupation of the southern cone. Current Anthropology, 48 164:174

Peres-Neto PR, and DA Jackson (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. Oecologia (Berl.). 129: 169–178

Ratto N (2003) Estrategias de caza y propiedades del registro arqueológico en la Puna de Chaschuil (Dpto de Tinogasta, Catamarca, Argentina). Doctoral dissertation. Facultad de Filosofía y Letras. Universidad de Buenos Aires. MS

Rohlf FJ (1990) Fitting curves to outlines. Proceedings of the Michigan morphometrics workshop. Special Publication No. 2, University of Michigan Museum of Zoology, AnnArbor 167:178

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. Contributions to Morphometrics, Edited by LF Marcus, E Bello and A Garcia-Valdecasas, Museo Nacional de Ciencias Naturales (CSIC). Madrid. Spain. 8 131:159

Rohlf FJ (1998) Review of "Fourier descriptors and their applications in biology". Bull. Math. Biol. 60:604–605.

Rohlf FJ (1999) Shape statistics: procrustes superimpositions and tangent spaces. Journal of Classification 16, 197:223

Rohlf FJ and M Corti (2000) The use of partial least-squares to study covariation in shape. Systematic Biology 49 740:753

Rohlf FJ (2002a) tps software series. http//life.bio.sunysb.edu/morph/

Rohlf FJ (2002b) Geometric morphometrics in phylogeny. In Morphology, shape and phylogenetics. Edited by P Forey and N Francis and Taylor. London. 175:193

Rohlf FJ and FL Bookstein (2003) Computing the uniform component of shape variation. Systematic Biology 52, 66:69

Saitou N and M Nei (1987) The neigbor-joining method: a new metod for reconstructing pylogenetic trees. Molecular Biology and Evolution 4(4) 406:425

Scartascini F and M Cardillo (2009) Explorando la variabilidad métrica y morfológica de las "pesas líticas" recuperadas en el sector norte de la costa del golfo de San Matías. Arqueometría Latinoamericana. 2do. Congreso Argentino y 1ro Latinoamericano. Vol. 1, 162–175. Buenos Aires. Argentina

Saragusti I, I Sharon, O Katzenelson and D Avnir (1998) Quantitative analysis of the symmetry of artifacts: lower Paleolithic handaxes. Journal of Archeological Science 25, 817:825

Saragusti I, A Karasik, I Sharon and U Smilansky (2005) Quantitative analysis of shape attributes based on contours and section profiles in artifact analysis. Journal of Archaeological Science 3 841:853

Sheets HD (2003) IMP-Integrated Morphometrics Package. Department of Physics, Canisius College, Buffalo, New York. http://www2.canisius.edu/~sheets/morphsoft.html

Sheets HD, K Kim and CE Mitchell (2004) A combined landmark and outline-based approach to ontogenetic shape change in the Ordovician trilobite Triathrus becki. Morphometrics applications in biology and paleontology. Edited by AMT Elewa. Berlin. Springer. 67:82

Sheets HD, KM Covino, JM Panasiewicz and SR Morris (2006) Comparison of geometric morphometric outline methods in the discrimination of age-related differences in feather shape. Frontiers in Zoology 3:15

Shott MJ (1996) Innovation and selection in prehistory. A case study from the American bottom. Stone tools theoretical insights into human prehistory. Edited by G Odell, Plenum Press. New York. 279:309

Shott MJ, DA Hunzicker and B Patten (2007) Pattern and allometric measurement of reduction in experimental Folsom bifaces. Lithic Technology 32(2) 203:217

Slice DE (1998) Morpheus et al. Software for morphometric research department of ecology and evolution, State University of New York. Stony Brook. NY. http://www.morphometrics.org/morpheus.html

Wynn T and F Tierson (1990) Regional comparison of the shapes of later Acheulian handaxes. American Anthropologist 92, 73:87

Zelditch ML, DL Swiderski, DH Sheets and WL Fink (2004) Geometric Morphometrics for Biologist. A Primer. Elsevier. Academic Press. NY

Part IV
Scope to the Future of Morphometrics
(Chaps. 16 to 18)

Chapter 16
Prospectus: The Future of Morphometrics

Richard E. Strauss

Idea and Aims

The field of morphometrics has transitioned relatively smoothly through several different phases, from D'Arcy Thompson's (1917) extraordinary and influential treatise on growth and form, through the influx of algebraic and statistical methods related to eigenanalysis, cluster analysis, and multidimensional scaling, to direct landmark-based Procrustes and deformation methods that echo Thompson's original intents and insights. In a sense, the discipline is still riding the wave of methodological advances that began in the 1970s (Adams et al. 2004). Although it is difficult to predict to direction of future methodological advances, it is certain that morphometric methods will be extended to areas currently at the periphery of current applications, such as the use of morphometrics to study the effects of quantitative trait loci (Klingenberg et al. 2001; Klingenberg 2003; Leamy et al. 2008) and the sizes and shapes of molecules (Billoud et al. 2000; Bookstein 2004; Rogen and Bohr 2003). However, even given the current level of methodological sophistication, there are still some important technical and conceptual problems to be solved in the shorter term. I will briefly highlight just a few of these here.

Three-Dimensional Analyses

The extension from 2-dimensional to 3-dimensional analyses has been available in principle for many years, and the use of 3D landmarks is becoming standard practice in fields such as physical anthropology and biomedical science. Procrustes methods are easily extended to three (or more) dimensions, and analyses of Procrustes residuals (Berge and Penin 2004; Lockwood et al. 2002; Nicholson and Harvati 2006) and, more recently, application of three-dimensional extensions of thin-plate splines and other morphometric methods are becoming more widely used (Bookstein 1989;

R.E. Strauss (✉)
Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409-3131, USA
e-mail: Rich.Strauss@ttu.edu

(A) Reference form   (B) Affine component

(C) Non-affine component   (D) Total transformation
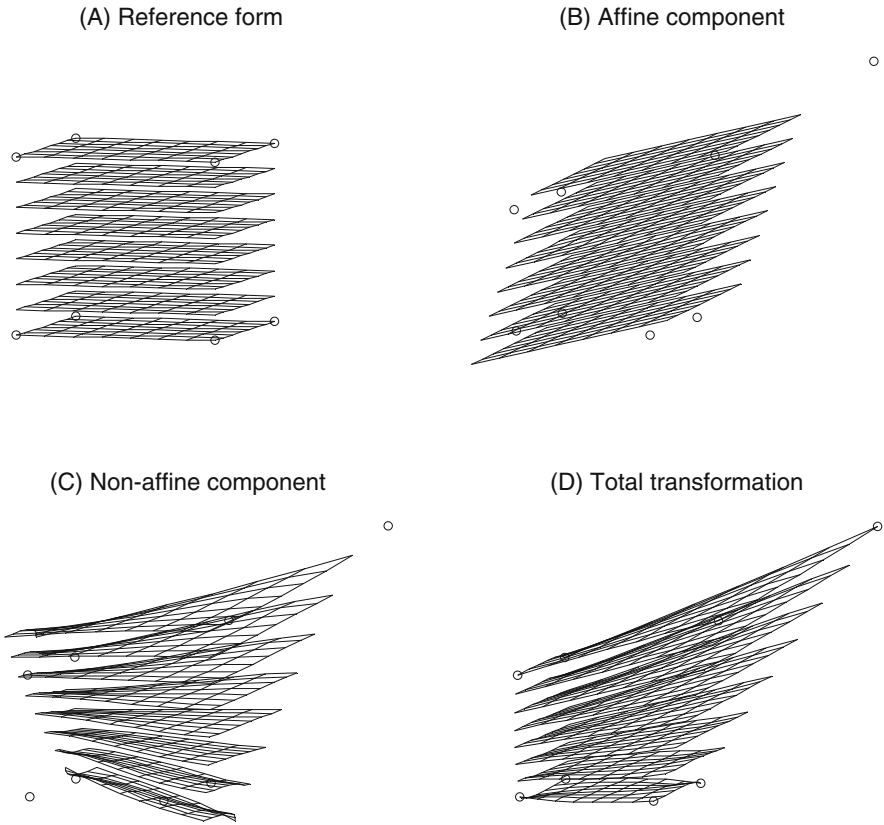
Fig. 16.1 Deformation of a reference form (a cube) to a target form based on the 3D thin-hyperplate spline interpolation. In the target form two landmarks (*top right*) have been displaced, while the others remain in place. The total transformation fits the points exactly, while the affine and non-affine components represent, respectively, global and local components of the total transformation

Gunz et al. 2005; Mitteroecker et al. 2004; Mitteroecker and Gunz 2002; O'Higgins and Jones 1998; Rohlf and Bookstein 2003; Rohr 2001).

Because illustrations of 3D thin-plate splines are rare, a simple example is illustrated in Fig. 16.1. This "thin-hyperplate" spline corresponds to the distortion of a malleable cube. The shift of any arbitrary point is given by a weighted sum of landmark shifts, in which each landmark shift is weighted according to its distance from the point (Bookstein 1989). The weights for interpolated shifts of arbitrary points in three dimensions are $-U(d) = -|d|$, where d is the distance from the point to a landmark. The stacked 2D planar grids for the reference form are a sample of slices through what is actually a solid 3D grid, but the planar grids are useful for visualization. The corresponding layers in the final deformed "cube" are, of course, not necessarily planar even for subsets of landmarks that lie in a plane in both the

reference and target forms. Although, in parallel to the 2D case, the non-affine component of the spline can be decomposed into eigenfunctions in the 3D case, there are a number of computational details that need to be resolved. Bookstein (1989) listed several aspects of the 3D extension that "will require considerable imagination". In particular, since 2D splines are visualized in 3D, it is not clear how best to draw a thin-hyperplane spline that is a 3D projection of a 4D geometric object.

Landmark Variation

The problem of how to adequately characterize relative amounts and directions of variation at individual landmarks is of interest to many researchers, particularly those utilizing Procrustes analyses of 3D coordinates. Results depend sensitively on the particular method used to superimpose configurations of landmarks (the so-called registration problem). Although several versions of Procrustes alignment are available and their statistical properties have been well characterized (Rohlf 1990; Rohlf and Slice 1990), there are no apparent biological rationales for choosing among them. Minimizing the sum-of-squared deviations among corresponding landmarks on different forms tends to produce circular distributions of coordinate positions, with approximately equal amounts of dispersion for different landmarks, while minimizing the median deviations tends to produce more elliptical distributions and more variation among landmarks. Currently these statistical patterns cannot be distinguished from inherent biological variability.

Allometry

More direct ties to developmental biology are needed (Gilbert 2003; Klingenberg 2002), particularly with respect to allometric scaling. The use of power functions to characterize scaling relationships (Huxley 1932) has a long and venerable history (Brown et al. 2000; Strauss 1993), and in most biological disciplines the term "allometry" is virtually synonymous with use of power functions (and, more recently, with fractal dimensions). In the context of geometric morphometrics, however, shape is defined much more generally as the "geometric information that remains when location, scale and rotational effects are filtered out from an object" (Bookstein 1978; Kendall 1977, etc.). The Procrustes method is used in geometric morphometrics to standardize forms to a common centroid size, which represents an isometric size standardization (Bookstein 1991; Rohlf 1993). Consequently, the concept of allometry has been generalized to *any* variation in shape that is correlated with size (Bookstein 1991; Zelditch et al. 2003). The null hypothesis both for the geometric model and for Huxley's model is isometry; however, deviation from the null in Huxley's model represents a particularly constrained form of anisometry. In geometric morphometric analyses the coefficients describing shape differences are not meaningful in terms of specific growth models such as Huxley's power law (Zelditch et al. 2004). Whereas deformation models are portrayed in the linear

space, Huxley's model is linear in the log-space. The anisometric "shape" variation studied in geometric morphometrics therefore consists of two components: allometric (*sensu* Huxley) and non-allometric. The issue of allometric size-adjustment (*sensu* Huxley) of Procrustes residuals or of deformation grids needs to be pursued further (e.g., Hammer 2004).

Missing Data

Missing data are a frequent problem in morphometrics, as they are generally in multivariate statistical analyses (Reig 1998; Richtsmeier et al. 1992; Strauss et al. 2003; Strauss and Atanassov 2006; Yaroch 1996). If a form is distorted (as in fossils) or has been damaged or broken off (as in delicate skeletal samples), then landmarks can be missing in some specimens. In statistical studies there are two main strategies for dealing with missing data: either the variables with missing data (coordinates of a missing landmark, in this case) must be ignored, or missing values must be imputed (i.e., estimated from the values in complete specimens). Gunz et al. (2002) have summarized how knowledge about the context of missing landmarks can be used to approximate their positions. This would include factors such as bilateral symmetry (landmarks on one side of the body can be "mirrored" to the other side), allometry (regression of Procrustes shape coordinates on centroid size), morphological integration (quantifying patterns of covariation of subsets of landmarks), and curvature smoothness (as quantified by magnitudes of deformation associated with the thin-plate spline interpolation). However, only a few preliminary comparative studies of these different approaches have been carried out, and much additional work needs to be done to characterize the best strategies for dealing with missing data.

Phylogenetics

Because morphometric studies are often carried out within a phylogenetic context, the use of morphometric data in phylogenetic analyses continues to be a contentious topic. MacLeod (2002) discussed this general question and suggested that the hesitation expressed by many at the use of morphometric data in phylogenetics can be traced back to the strong historical connections between morphometrics and phenetics, which was formulated as a philosophy of systematics in the 1960s and 1970s (Pimentel 1979; Sneath and Sokal 1973) and later directly contrasted with the aims and methods of cladistics.

There are three main areas in which morphometrics can play a role: (a) in the generation of characters useful for phylogenetic inference (i.e., estimation of trees); (b) in the interpretation of morphological diversification within the context of a phylogenetic hypothesis produced with other data (typically molecular sequence data); and (c) in "ancestral reconstruction", the inference about morphological states in ancestors. Of these, the first has been the most problematic. Although many systematists argue that there is no inherent reason for disregarding

the use of morphometric data in phylogenetics (Guerrero et al. 2003; MacLeod 2002), others have disagreed (Cranston and Humphries 1988; Crowe 1994; Curnoe 2003; Pimentel and Riggins 1987). But even among those who agree, there is little consensus on the kinds of data that are most appropriate and how they are best to be used. Most relevant to geometric morphometrics have been the arguments by Zelditch et al. (1995), Swiderski (1993), Fink and Zelditch (1995) and Zelditch and Fink (1995) that particular morphometric methods are compatible with the concept of taxic homology. These authors claimed that latent variables such as partial or principal warp decompositions are the only morphometric variables applicable in phylogenetic contexts. Others (Bookstein 1994; Felsenstein 2002; Lynch et al. 1996; Naylor 1996; Rohlf 1998) have challenged or cautioned about the use of geometric morphometric variables for phylogenetic analysis. Additional enlightenment may come from application rather than theory. For example, González-José et al. (2008) have recently shown that when continuous, correlated, modularized morphometric characters are treated as such, cladistic analysis (which is conventionally based on discrete, hypothetically independent characters) can successfully resolve phylogenetic relationships among species of *Homo*.

MacLeod (2001) reviewed the basic principles involved in viewing morphometric variation within the context of phylogenetically structured comparisons. Excellent recent examples of the use of morphometrics to interpret patterns of morphological diversification include Guill et al. (2003), Magniez-Jannin et al. (2004), and Larson (2005). As for ancestral reconstruction, Rohlf (2002) described a method for estimating ancestral states of shape variables using "squared-change parsimony", and using the inferred states to depicting shape changes between nodes of the tree as deformations and to estimate the image of an ancestor.

Conclusion

The geometric morphometric methods that have been developed over the past several decades to extend beyond the limitations of traditional distance-based methods have become transformed into the new standard research protocol. As the technologies of measurement, analysis and display continue to improve, it will be interesting to see how the current methods evolve over the next few decades.

References

Adams DC, Rohlf FJ, Slice DE (2004) Geometric morphometrics: ten years of progress following the 'revolution'. Ital J Zool 71: 5–16.
Berge C, Penin X (2004) Ontogenetic allometry, heterochrony, and interspecific differences in the skull of African apes, using tridimensional Procrustes analysis. Am J Phys Anthropol 124: 124–138.

Billoud B, Guerrucci MA, Deutsch JS (2000) Cirripede phylogeny using a novel approach: molecular morphometrics. Mol Biol Evol 17: 1435–1445.

Bookstein FL (1978) The Measurement of Biological Shape and Shape Change. Lecture Notes in Biomathematics 24. Springer-Verlag, New York.

Bookstein FL (1989) Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans Pattern Anal Machine Intell 11: 567–585.

Bookstein FL (1991) Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, Cambridge.

Bookstein FL (1994) Can biometrical shape be a homologous character? In Hall BK (ed) Homology: The Hierarchical Basis of Comparative Biology. Academic Press, San Diego, pp 197–227.

Bookstein FL (2004) On a surprising bridge between morphometrics and bioinformatics. Bioinformatics, images, and wavelets, pp. 41–45. Leeds, UK, Leeds Annual Statistical Research Workshop.

Brown JH, West GB, Enquist BJ (2000) Scaling in biology: patterns and processes, causes and consequences. In Brown JH, West GB (eds) Scaling in Biology. Oxford University Press, New York, pp 1–24.

Cranston PS, Humphries CJ (1988) Cladistics and computers: a chironomid conundrum? Cladistics 4: 72–92.

Crowe TM (1994) Morphometrics, phylogenetic models and cladistics: means to an end or much to do about nothing? Cladistics 10: 77–84.

Curnoe D (2003) Problems with the use of cladistic analysis in palaeoanthropology. HOMO 53: 225–234.

Felsenstein J (2002) Quantitative characters, phylogenies, and morphometrics. In MacLeod N, Forey PL (eds) Morphology, Shape and Phylogeny. Taylor and Francis, London, pp 27–44.

Fink WL, Zelditch ML (1995) Phylogenetic analysis of ontogenetic shape transformations: a reassessment of the piranha genus *Pygocentrus* (Teleostei). Syst Biol 44: 343–360.

Gilbert SF (2003) The morphogenesis of evolutionary developmental biology. Int J Dev Biol 47: 467–477.

González-José R, Escapa I, Neves WA, Cúneo R, Pucciarelli HM (2008) Cladistic analysis of continuous modularized traits provides phylogenetic signals in *Homo* evolution. Nature 453: 775–778.

Guerrero JA, De Luna E, Sánchez-Hernández C (2003) Morphometrics in the quantification of character state identity for the assessment of primary homology: an analysis of character variation of the genus *Artibeus* (Chiroptera: Phyllostomidae). Biol J Linn Soc 80: 45–55.

Guill JM, Heins DC, Hood CS (2003) The effect of phylogeny on interspecific body shape variation in darters (Pisces: Percidae). Syst Biol 52: 488–500.

Gunz P, Mitteroecker P, Bookstein FL (2005) Semilandmarks in three dimensions. In Slice DE (ed) Modern Morphometrics in Physical Anthropology. Plenum Publishers, New York, pp 73–98.

Gunz P, Mitteroecker P, Bookstein FL, Weber GW (2002) Approaches to missing data in anthropology. Collegium Antropologicum 26: 78–79.

Hammer Ø (2004) Allometric field decomposition – an attempt at morphogenetic morphometrics. In Elewa AMT (ed) Morphometrics: Applications in Biology and Paleontology. Springer, Berlin, pp 55–65.

Huxley JS (1932) Problems of Relative Growth. Methuen, London.

Kendall DG (1977) The diffusion of shape. Adv Appl Prob 9: 428–430.

Klingenberg CP (2002) Morphometrics and the role of the phenotype in studies of the evolution of developmental mechanisms. Gene 287: 3–10.

Klingenberg CP (2003) Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach. Evolution 57: 191–195.

Klingenberg CP, Leamy LJ, Routman EJ, Cheverud JM (2001) Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics 157: 785–802.

Larson PM (2005) Ontogeny, phylogeny, and morphology in anuran larvae: morphometric analysis of cranial development and evolution in *Rana* tadpoles (Anura: Ranidae). J Morphol 264: 34–52.

Leamy LJ, Klingenberg CP, Sherratt E, Wolf JB, Cheverud JM (2008) A search for quantitative trait loci exhibiting imprinting effects on mouse mandible size and shape. Heredity 101: 518–526.

Lockwood CA, Lynch JM, Kimbel WH (2002) Quantifying temporal bone morphology of great apes and humans: an approach using geometric morphometrics. J Anat 201: 447–464.

Lynch JM, Wood CG, Luboga SA (1996) Geometric morphometrics in primatology: craniofacial variation in *Homo sapiens* and *Pan troglodytes*. Folia Primatol 67: 15–39.

MacLeod N (2001) The role of phylogeny in quantitative paleobiological data analysis. Paleobiology 27: 226–249.

MacLeod N (2002) Phylogenetic signals in morphometric data. In MacLeod N, Forey PL (eds) Morphology, Shape and Phylogeny. Taylor and Francis, London, pp 100–138.

Magniez-Jannin F, David B, Dommergues JL, Zhi-Hui S, Okada TS, Osawa S (2004) Analysing disparity by applying combined morphological and molecular approaches to French and Japanese carabid beetles. Biol J Linn Soc 71: 343–358.

Mitteroecker P, Gunz P (2002) Semilandmarks on curves and surfaces in three dimensions. Am J Phys Anthropol 34 (Suppl): 114–115.

Mitteroecker P, Gunz P, Bernhard M, Schaefer K, Bookstein FL (2004) Comparison of cranial ontogenetic trajectories among great apes and humans. J Hum Evol 46: 679–697.

Naylor GJP (1996) Can partial warp scores be used as cladistic characters? In Marcus LF, Corti M, Loy A, Naylor GJP, Slice DE (eds) Advances in Morphometrics. Plenum Press, New York, pp 519–530.

Nicholson E, Harvati K (2006) Quantitative analysis of human mandibular shape using three-dimensional geometric morphometrics. Am J Phys Anthropol 131: 368–383.

O'Higgins P, Jones N (1998) Facial growth in *Cercocebus torquatus*: an application of three-dimensional geometric morphometric techniques to the study of morphological variation. J Anat 193: 251–272.

Pimentel RA (1979) Morphometrics: the Multivariate Analysis of Biological Data. Kendall-Hunt, Dubuque.

Pimentel RA, Riggins R (1987) The nature of cladistic data. Cladistics 3: 201–209.

Reig S (1998) 3D digitizing precision and sources of error in the geometric analysis of weasel skulls. Acta Zoologica Academiae Scientiarum Hungaricae 44: 61–72.

Richtsmeier JT, Cheverud JM, Lele SR (1992) Advances in anthropological morphometrics. Annu Rev Anthropol 21: 283–305.

Rogen P, Bohr H (2003) A new family of global protein shape descriptors. Math Biosci 182: 167–181.

Rohlf FJ (1990) Rotational fit (Procrustes) methods. In Rohlf FJ, Bookstein FL (eds) Proceedings of the Michigan Morphometrics Workshop. University of Michigan Museum of Zoology, Ann Arbor, pp 227–236.

Rohlf FJ (1993) Relative warp analysis and an example of its application to mosquito wings. In Marcus LF, Bello E, Garcia-Valdecasas A (eds) Contributions to Morphometrics. Museo Nacional de Ciencias Naturales, Madrid, Spain, pp 131–159.

Rohlf FJ (1998) On applications of geometric morphometrics to studies of ontogeny and phylogeny. Syst Biol 47: 147–158.

Rohlf FJ (2002) Geometric morphometrics and phylogeny. In MacLeod N, Forey PL (eds) Morphology, Shape and Phylogenetics. Taylor and Francis, London, pp 175–193.

Rohlf FJ, Bookstein FL (2003) Computing the uniform component of shape variation. Syst Biol 52: 66–69.

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superposition of landmarks. Syst Zool 39: 40–59.

Rohr K (2001) Landmark-Based Image Analysis: Using Geometric and Intensity Models. Kluwer, Norwell.

Sneath PHA, Sokal RR (1973) Numerical Taxonomy: the Principles and Practice of Numerical Classification. W.H. Freeman, San Francisco.

Strauss RE (1993) The study of allometry since Huxley. In Huxley JS (ed) Problems of Relative Growth. Johns Hopkins University Press, Baltimore, pp 47–75.

Strauss RE, Atanassov MN (2006) Determining best subsets of specimens and characters in the presence of large amounts of missing data. Biol J Linn Soc 88: 309–328.

Strauss RE, Atanassov MN, Oliveira JA (2003) Evaluation of the principal-component and expectation-maximization methods for estimating missing data in morphometric studies. J Vert Paleontol 23: 284–296.

Swiderski DL (1993) Morphological evolution of the scapula in tree squirrels, chipmunks and ground squirrels (Sciuridae): an analysis using thin-plate splines. Evolution 47: 1854–1873.

Thompson DW (1917) On Growth and Form (Reprinted From 1942 Edition). Dover, New York.

Yaroch LA (1996) Shape analysis using the thin-plate spline: neanderthal cranial shape as an example. Yearb Phys Anthropol 39: 43–89.

Zelditch ML, Fink WL (1995) Allometry and developmental integration of body growth in a piranha, *Pygocentrus nattereri* (Teleostei: Ostariophysi). J Morphol 223: 341–355.

Zelditch ML, Fink WL, Swiderski DL (1995) Morphometrics, homology, and phylogenetics: quantified characters as synapomorphies. Syst Biol 44: 179–189.

Zelditch ML, Lundrigan BL, Sheets HD, Garland T, Jr. (2003) Do precocial mammals develop at a faster rate? A comparison of rates of skull development in *Sigmodon fulviventer* and *Mus musculus domesticus*. J Evol Biol 16: 708–720.

Zelditch ML, Swiderski DL, Sheets HD, Fink WL (2004) Geometric Morphometrics for Biologists: A Primer. Academic Press, New York.

Chapter 17
Morphometrics and Cosmology: Short Note and Future Hope

Ashraf M.T. Elewa

"By using a D'Arcy machine to begin a study of microbial life on Earth, someday remote and automated instruments may be able to identify life elsewhere in the universe – whatever form that life may take". This paragraph is mentioned in an interesting article titled "Who Wrote the Book of Life?" written by Leslie Mullen (6 Jan 2001) on the firstscience.com website. It is also mentioned in MARSBUGS, The Electronic Astrobiology Newsletter, before that date (28 May 1999).

Is this true? Is there life outside Earth? Were there creatures elsewhere in the universe? Are there creatures living, for example, on Mars? These questions and more arise to mind when you think of this mystery. Several scientists devoted their research to search for life outside Earth (e.g. Goldsmith and Owen 1992; Davies 1995; Sagan 1995; Angel and Woolf 1996; Goldsmith 1997; Walter 1999; Scott 2008).

Firstly, to clarify this topic it is important to introduce the David Harland's thought in his book titled "Water and the Search for Life on Mars", which is published in the year 2005. He believes that searching for life on Mars should be associated with studying the origin of life on Earth. This means that if we could discover, from the rocks of Mars, what is believed to be the earliest form of life (cyanobacteria with chlorophyll for photosynthesis), then we can provide evidence that the same could be the case on Mars.

Consequently, we should think of three possibilities (hypotheses) to answer the above mentioned questions:

1. There are no and there were no life and creatures on Mars; or
2. There were life and creatures on Mars; or
3. There are life and creatures on Mars.

The oldest and most famous evidence of life possibility on Mars is the meteorite ALH84001, which was discovered in Allan Hills of Antarctica in the year 1984.

A.M.T. Elewa (✉)
Geology Department, Faculty of Science, Minia University, Box 61519, Minia, Egypt
e-mail: aelewa@link.net; ashrafelewa@ymail.com

The idea could be proved if we can establish somesort of paleoenvironmental similarity between Earth and Mars. The ideal example is represented by the thermal springs that were associated with extreme environments. Astrobiologists believe in the likelihood of abundant thermal spring environments on early Earth and Mars.

Tang and Roopnarine (2003) stated that thermal springs in evaporitic environments provide a unique biological laboratory in which to study natural selection and evolutionary diversification. These isolated systems may be an analogue for conditions in early Earth or Mars history.

Usually, cosmologists speak about and describe how meteorites strike Earth, and explain the results of the collision between these meteorites and the Earth. One of these possible severe results is the extinction of dinosaurs since about 65 million year. Many authors assigned this extinction to the extraterrestrial impact. Conversely, recent studies, including my own research, indicate multiple causes for this extinction. The most important part, however, is the age dating using radioactive elements. Some authors believe that $\Delta^{14}C$ has a half-life of 5750 years. Here I would note that some other scientists found it inaccurate to apply this technique for organic matters of more than 3,000 years old. I would also add that there should be definite cautions when applying this technique because any disturbance of the optimum conditions will normally lead to inaccurate results.

Dealing with our focus on the material suitable for morphometric analyses (carbonate globules), authors mentioned two significantly controversial ages to the carbonate globules present in the meteorite ALH84001. These conflicting results ensure my previously mentioned note on how the change in the surrounding conditions and the chemical composition affects the results of age dating.

I should note that all experiments (e.g. labeled-release, gas-exchange, and pyrolitic-release) made to verify life on Mars have promptly failed. Therefore, there should be another solution to solve the problem of discovering the origin of these carbonate globules.

Anyway, these globules contain aromatic hydrocarbons, magnetic minerals (iron oxides and sulphides), and bacteria-like forms but with significant smaller size than any bacteria discovered on Earth (for details see Goldsmith 1997).

Again, we should think of two hypotheses:

1. These carbonate globules that were formed within the fractures of the meteorite are from Mars; or
2. These globules were formed in the fractures of the meteorite after its collapse and collision with Earth.

It is possible that severe environmental conditions may lead to assemble these globules with their organic constituents in the meteorites after their falling down to Earth!! Solving this problem may help discovering the truth.

It seems that this solution could be found through studying shapes (morphometrics) of these bacteria-like forms in the carbonate globules. Though, how can we use morphometrics to do that?

I suggest the following scenario for that solution:

1. At first, we try to prove the hypothesis of significant variation between shapes of these bacteria-like forms and their resemblances of Earth origin that live under extreme conditions;
2. If this variation is significant, then these bacteria-like forms could be another type of bacteria characteristic to Mars or they were not living organisms at all;
3. To establish that these forms were organisms it is important to search for characteristics of living forms in the carbonate globules (e.g. evidences of cell division, the walls of these cells were made of organic matter . . . etc.). However, it is not easy to do that with such very old rock. If we substantiated that these forms were organisms, then we should try to discover their origin (Earth or Mars) in the next step:
4. If the variation of shapes is insignificant, we may conclude probable affinity connecting these two groups; consequently
5. We study the environmental conditions under which the bacteria from Earth could live. At this point, the close similarity between the environmental conditions of the two studied groups supports the second hypothesis that is mentioned above. In contrast, dissimilarity confirms the first hypothesis.

Even though, one may ask how can we study the environmental conditions of the bacteria-like forms. The answer is very simple; you can analyze the carbonate globules using isotope analyses to define elements leading to environmental significances.

As a final point, I did not discuss the morphometric tools and techniques that are relevant to analyze such data, it is not my mission here to review them; instead, I gave notice to one of several scientific branches that could be developed using morphometrics as a stride towards better future.

References

Angel R, Woolf N (1996) Searching for life on other planets. Scientific American 274 : 60–66

Davies P (1995) Are we alone? New York, Basic Books

Goldsmith D (1997) The hunt for life on Mars. Penguin, USA

Goldsmith D, Owen T (1992) The search for life in the universe (2nd ed).Addison-Wesley, Reading

Harland DM (2005) Water and the search for life on Mars. Springer Praxis Books / Space Exploration, Berlin

MARSBUGS (1999) The Electronic Astrobiology Newsletter 6 (14)

Sagan C (1995) Pale blue dot: A vision of the human future in space. Headline Book Publishing, London

Scott E (2008) Mars and the search for life. Clarion Books, New York

Tang CM and Roopnarine PD (2003) Evaporites, water, and life, Part I: Complex morphological variability in complex evaporitic systems – Thermal spring snails from the Chihuahuan Desert, Mexico. Astrobiology 3 (3): 597–607

Walter M (1999) Search for life on Mars. New York, Basic Books

Chapter 18
Morphometrics in Past: Integrating Morphometrics with General Data Analysis Software

Øyvind Hammer

Idea and Aims

The use of the software package Past for morphometrics is described. The advantage of integrating morphometrics with both general statistical and special ecological analysis modules in the same program is emphasized. A spreadsheet-type user interface is easy to use but also allows flexibility in the chaining of analyses. This combination of features makes Past especially convenient for teaching.

The Past Software

A morphometric study is usually part of a larger, interdisciplinary project, involving e.g. environmental, ecological or genetic data. It may therefore be advantageous to integrate morphometric and other statistical methods in the same software package, to reduce data conversion and the learning of different user interfaces. The R Project (R Development Core Team 2008) is presently without doubt the most wide-ranging and comprehensive software in this respect, including tools for both geometric morphometrics, general data analysis and specialized analysis within other fields. It is also highly flexible. However, this flexibility comes at the cost of a steep learning curve. On the other hand, most of the available morphometrics packages, such as the TPS range (Rohlf), MorphoJ (Klingenberg 2008), Morphologika (O'Higgins and Jones 2006), Morpheus et al. (Slice 2008) and IMP (Zelditch et al. 2004), are user friendly and with a smooth workflow for geometric morphometrics, but may need to be used in conjunction with other software if non-morphometric analysis methods are part of the study.

The free software package Past (Paleontological Statistics) for Windows (Hammer et al. 2001) occupies a middle ground along this gradient. Past is a general statistics package but also with a range of specialized modules for analysis

Ø. Hammer (✉)
Natural History Museum, University of Oslo, Oslo, Norway
e-mail: oyvind.hammer@nhm.uio.no

within ecology, systematics, stratigraphy, spatial and time series analysis. Basic morphometric modules include 2D and 3D Procrustes alignment, thin-plate spline deformation grids, partial and relative warps, and outline analysis. The results from these operations can be used directly as input to e.g. PCA, discriminant analysis, MANOVA (including several classes of permutation test), PLS, cluster analysis, Mantel tests and a range of regressions and classical univariate statistics and tests.

Past is presently at version 1.88. It is currently cited in around 300 scientific publications annually, and is downloaded around 50,000 times per year. This large user base means that most functions are well tested. The program is based on a standard Excel-like spreadsheet, and can read files in e.g. Excel, tab-separated text, TPS and Nexus formats. Output is presented both graphically and numerically, with many plotting options and in publication-quality vector format. Although the program is primarily designed for interactive data exploration, a full scripting language is also included.

The Morphometrics Workflow in PAST

The open philosophy of the software will be illustrated with two simplistic examples.

Morphometrics and Genetics of Hominids

Skulls representing four extant hominid genera (*Homo*, *Gorilla*, *Pan*, *Pongo*) and one New-World monkey (*Saimiri*) were selected for analysis. Ten landmarks were digitized from lateral views of CT scans available from the Digital Morphology Library at the University of Texas at Austin. Mitochondrial DNA sequence data were taken from Hayasaka et al. (1988). The morphometric data were saved from the tpsDig program (Rohlf) in the TPS format, while the sequence data are available in the Nexus format.

The two files can be opened directly in Past. In this example, the morphometric and the genetic data sets are placed into the same spreadsheet by the user. The landmark coordinates are then subjected to Procrustes fitting (Fig. 18.1) and Principal Components Analysis (Fig. 18.2). The bootstrapping analysis in the PCA module indicates that only the first component is robust under resampling. This first relative warp is shown as a deformation grid in Fig. 18.2 (lollipop plots are also available), showing that positive scores on PC1 correspond to a compression of the posterior relatively to the anterior part of the skull. The scatter plot shows a gradient in PC1 from *Homo* through *Saimiri* to *Pan*, *Pongo* and *Gorilla*.

The genetic data set can be subjected to independent analysis. Figure 18.3 shows two analyses based on the Jukes–Cantor distance measure for genetic sequence data. The neighbour joining clustering diagram can be interpreted as a simple phylogenetic tree with very high bootstrap values (cladistic parsimony analysis is also available). The NMDS ordination has a stress of 0.0.
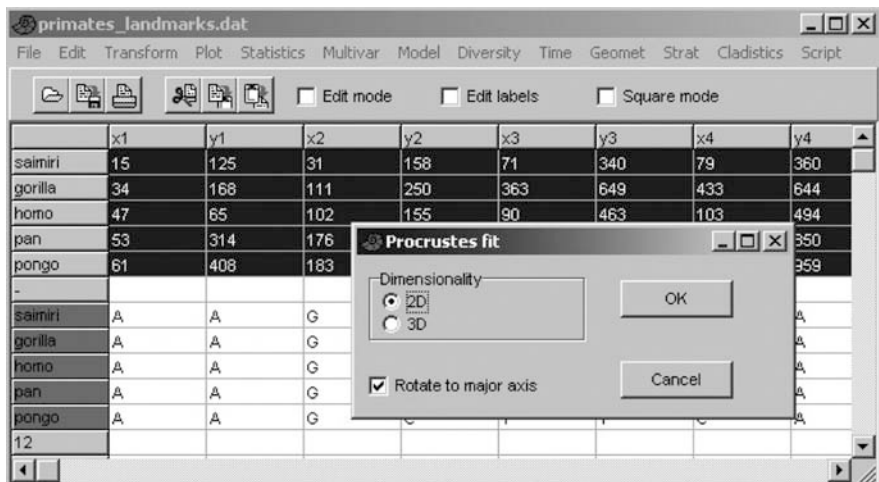
Fig. 18.1 The Past interface. The user has loaded the landmark coordinate data and the sequence data into the same spreadsheet, selected the coordinate data and is performing a Procrustes fit
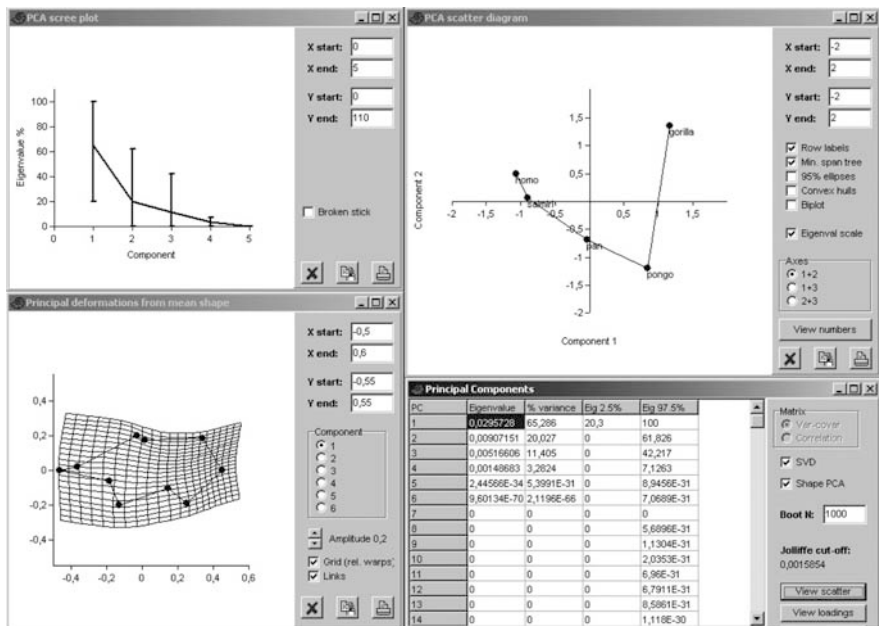
Fig. 18.2 Some of the windows in the PCA module. *Upper left*: Scree plot with bootstrapped 95% confidence intervals for the eigenvalues. *Upper right*: PCA scatter plot with minimal spanning tree. *Lower right*: PCA master window. *Lower left*: First principal component visualized as a deformation grid
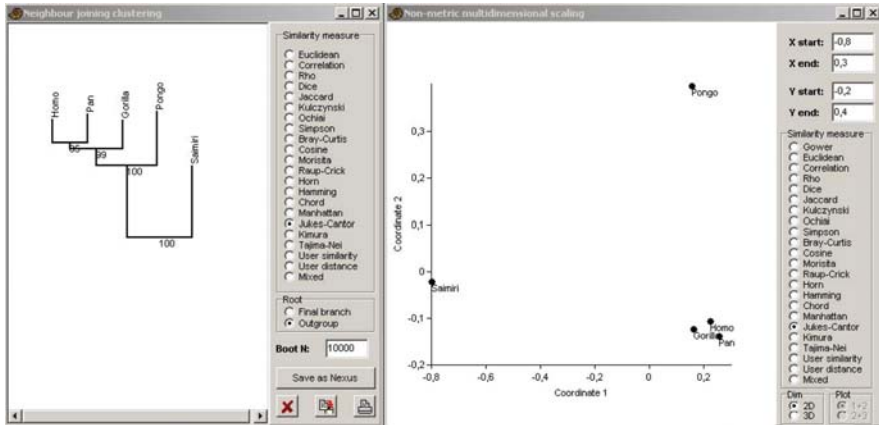
Fig. 18.3 Neighbour joining clustering (*left*) rooted on *Saimiri*, and non-metric multidimensional scaling (*right*), both based on the Jukes–Cantor distance measure for genetic sequence data

These genetic analyses are in correspondence with standard phylogenetic hypotheses for hominids, but the analyses based on morphometrics are not. This statement can be quantified by Past using a Mantel test for correlation between the morphometric and genetic distance matrices. We use Euclidean distance between Procrustes coordinates (Procrustes distance) and the Jukes–Cantor distance between sequences. The correlation value $R = -0.187$, with a significance for no correlation at $p = 0.736$. In other words, there is a weak and nonsignificant negative correlation between morphometric and genetic distance. For this data set, morphology does not reflect phylogeny.

Allometry in Trilobites

The second example uses a dataset with digitized outlines of 51 cephala of the trilobite *Trimerocephalus* (Crônier et al. 1998). The cephala are divided into four instars (growth stages). Figure 18.4 shows the outlines in the upper left panel. The outlines were subjected to Elliptic Fourier Analysis. The EFA coefficients were copied back to the spreadsheet and subjected to PCA. The PCA scores were again copied back to the spreadsheet and used as shape descriptors for subsequent analysis. The lower right panels in Fig. 18.4 shows a MANOVA carried out on the four groups, indicating a highly significant overall difference in shape. The pairwise post-hoc tests show that instars 1 versus 2 and instars 3 versus 4 are not significantly different.

Centroid size was calculated from the original data, and the principal component scores regressed with size as independent variable. This multiple regression is also significant (lower left panel, Fig. 18.4), with PC2 showing the strongest correlation with size (linear regression shown).
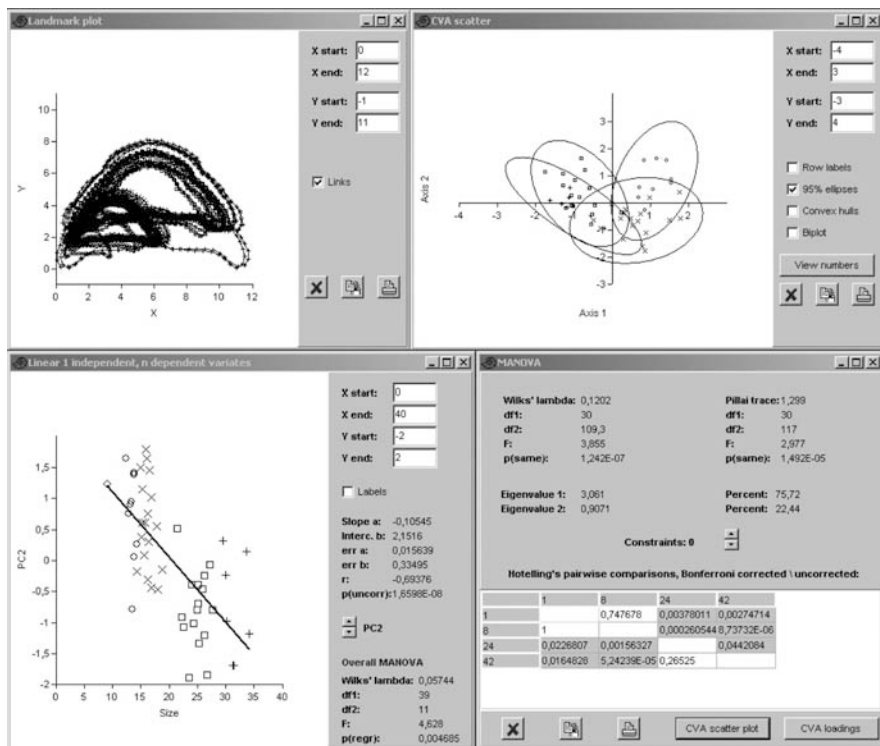
Fig. 18.4 *Upper left*: Digitized outlines of 51 trilobite cephala grouped into four instars. *Lower right*: MANOVA of the difference in shape as quantified by Elliptic Fourier Analysis coefficients. *Upper right*: Canonical variates analysis (discriminant analysis) of the four instars. *Lower left*: Multiple regression of shape onto size, here plotting size versus PC2 scores

Conclusions

In an ideal world, biologists should also be programmers and statisticians, but this is not generally the case and the difficulty in using advanced tools such as R may then become an obstacle. Both for education and professional data analysis, there is a niche for software that encourages quick and simple data mining while still supporting a wide range of methods that can be combined in new ways. The combination of morphometrics with a range of general univariate and multivariate methods and also special methods used in ecology, genetics, stratigraphy and spatial and time series analysis encourages experimentation with multidisciplinary data sets.

References

Crônier, C., Renaud, S., Feist, R. & Auffray, J.-C. 1998. Ontogeny of *Trimerocephalus lelievrei* (Trilobita, Phacopida), a representative of the Late Devonian phacopine paedomorphocline: a morphometric approach. Paleobiology 24:359–370.

Hammer, Ø., Harper, D.A.T. & Ryan, P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4(1):99.

Hayasaka, K., Gojobori, T. & Horai, S. 1988. Molecular phylogeny and evolution of primate mitochondrial DNA. Molecular Biology and Evolution 5:626–644.

Klingenberg, C.P. 2008. MorphoJ. Faculty of Life Sciences, University of Manchester, UK. http://www.flywings.org.uk/MorphoJ_page.htm

O'Higgins, P. & Jones, N. 2006. Tools for statistical shape analysis. Hull York Medical School. http://hyms.fme.googlepages.com/resources

R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Slice, D. 2008. Morpheus et al.

Zelditch, M.L., Swiderski,D.L., Sheets, H.D. & Fink, W.L. 2004. Geometric morphometrics for biologists: a primer. Elsevier, London.

Index