

Morphometrics

– brief notes

Oyvind Hammer
Paläontologisches Institut und Museum, Zürich
ohammer@nhm.uio.no

Zürich, September 29, 2002

Contents

1	Introduction	2
2	Acquisition of morphometric data	4
3	Univariate hypothesis testing	8
4	Bivariate measurements, curve fitting and allometry	13
5	Asymmetry	19
6	Cluster analysis	21
7	Principal components analysis	23
8	Multivariate hypothesis testing and discriminant analysis	28
9	Methods for some special types of morphometric data	31
10	Outline analysis	33
11	Landmark-based methods	38
12	Thin-plate splines and warps	43
	Bibliography	49

Chapter 1

Introduction

Palaeontology is becoming a quantitative subject, like other sciences such as biology and geology. The demands are increasing on palaeontologists to support their conclusions using statistics and large data bases. Palaeontology is also moving in the direction of becoming more *analytical*, in the sense that fossil material is used to answer questions about environment, evolution and ecology. A quick survey of recent issues of journals such as *Paleobiology* or *Lethaia* will show this very clearly indeed.

In addition to hypothesis testing, quantitative analysis serves other important purposes. One of them is sometimes referred to as 'fishing' (or data mining), that is searching for unknown patterns in the data that may give us new ideas. Finally, quantitative methods will often indicate that the material is not sufficient, and can give information about where we should collect more data.

This short text will describe some methods for quantitative treatment of measurements on fossil or Recent organisms. In practice, this means using computers. The goal of such analysis may be to assess intraspecific variation in order to define or merge species, or to investigate ontogeny, allometry, microevolution, ecophenotypic effects, sexual dimorphism or asymmetry.

The methodology of morphometrics has gone through a revolution during the last twenty years, with a large number of journals, books and articles devoted to the subject. The foundation of morphometrics is still the 'old' statistical methods based on measurements of a single variable such as length or width. Solid and easy-to-understand statistical workhorses such as the t test are often preferable over complex, modern methods which may just obscure matters. But often we want to study complex variation in complicated geometries. It may then be necessary to try to reduce the complex shapes to just a few variables in order to simplify the problem. New methods, in particular those based on the study of landmarks, have made it easier to parameterize shape in this way, to visualize shape change and test hypotheses statistically.

In this little course we can not go in any depth about the mathematical and statistical basis for the methods - the aim is to give a practical overview. But if such methods are used in work that is to be published, it is important that you know more about the underlying assumptions and the possible pitfalls. For such information I can only refer to the literature.

Software

In this course we will use a simple, but relatively comprehensive computer program for Windows, called PAST (PALaeontological STatistics). This program has been designed for teaching, and contains a collection of data analysis methods that are commonly used by palaeontologists in a user-friendly package. It can also be used for 'real' work. PAST is free, and can be downloaded to your own computer from the following web page:

<http://folk.uio.no/ohammer/past/>

Here you will also find the manual for the program, and a number of 'case studies' demonstrating the different methods. Some of these examples will be used in the course.

For more free morphometrics software, including excellent programs such as Morpheus and the TPS suite, go to the web site at Stony Brook:

<http://life.bio.sunysb.edu/morph/>

Chapter 2

Acquisition of morphometric data

In addition to the good old tools such as rulers, calipers and goniometers there is a wide spectrum of more advanced technology available, making measurements on fossil specimens easier and more accurate. In addition, for landmark and outline analysis (see later chapters) it is normally not sufficient to measure distances or angles. Instead, we need to digitize points or lines in terms of their absolute coordinates in the plane or in space (x-y or x-y-z). For this we need special equipment.

Acquisition of 2D coordinates

For analysis of 2D data, there is one cheap, accurate and practical method which is recommended for digitizing the x-y coordinates: Take a photograph of the specimen, scan the photograph and mark the wanted points or lines on the computer screen with the mouse. The scanning step can of course be skipped using a digital camera. Photoshop and many other image processing programs can then give you the coordinates, but sometimes they are just given in pixels and you may want to convert to a more appropriate size measure depending on the image magnification and resolution. However, for many purposes we are only interested in relative numbers, and the unit does not matter.

The tpsDig program by Rohlf (download from the Stony Brook web site) has been made specifically for taking x-y-coordinates from images, and its data files can be read directly into PAST.

A concern with this method is the limited accuracy due to image resolution: All coordinates will be rounded to the nearest pixel. It is therefore important to carry out the procedure using the maximal resolution allowed by the equipment.

In some cases, landmarks can be difficult to see on a photograph. Points and lines can then be digitized directly on the specimen using a coordinate measuring machine as described below.

Acquisition of 3D coordinates

The direct acquisition of 3D coordinates can only be done with more specialized and expensive equipment. One option is a 3D laser scanner or a computer tomograph, but such scanners are expensive, sometimes inaccurate, and considerable manual work is normally needed in front of the computer in order to get the landmarks from the 3D-representation on the screen. Also, the points that are digitized on the surface of the specimen will rarely coincide with the chosen landmarks. The positions of landmarks

must therefore be found by interpolation, which reduces accuracy further. Finally, landmarks are often based on features which do not show up in the 3D model, for example being points on sutures between bones which do not give any 3D relief. The only possibility then is to use a laser scanner with a camera, mapping the surface colors onto the 3D model, but such equipment is prohibitively expensive.

A faster and more accurate method is to use a coordinate measuring machine. This is a mechanical instrument with a needle that is pointed onto the landmark, and sensors report the xyz position to a computer. The needle can also be traced along a line or brushed over the specimen in order to digitize the complete surface. Several different types exist, but the most common configuration is an arm with precise angular sensors in the joints. Such arms are made by several companies, including Microscribe (cheap but inaccurate) and FARO (expensive and accurate). The top-end FARO arms have a precision of 25 micrometers over a cubic metre volume (fig. 2.1). These arms can of course also be used in 2D, for flat specimens.



Figure 2.1: Coordinate measuring machine, arm type (FARO).

However, most coordinate machines are not appropriate for small specimens, say less than 2 centimeters. Only a few, specialized and expensive machines have been constructed for use under a microscope.

In Zürich we have a FARO arm with 50 micrometers accuracy, suitable for specimens from 2-100 cm across. It is easy to use, and produces data files that can be read directly into PAST.

Using the FARO arm

The FARO arm is in principle very easy to use, but it is necessary to learn to maneuver it smoothly. This motor skill, which is difficult to explain in words, takes less than an hour to acquire.

The operation of the arm proceeds as follows:

- Start up the computer and log in. User name and password will be given in the class.
- Switch on the FARO arm. The switch is mounted on the power cable. The arm should say beep-beep.
- After a power-up, you must move all joints of the arm around until all red lamps have turned off (one lamp will remain on when the arm is in resting position). This can be a little tricky, so don't switch the arm off unless you leave it for a long time.
- For simple measuring of landmarks, distances or contours, use the "Caliper 3D" program in the "Faro" submenu of the "Programme" menu. For details on operating this program, see the manual, but the basic operations are very simple.
- If you want, you can set the coordinate system using "Three points" in the "Alignments" menu, but for most applications in morphometry this is not really necessary.
- Now digitize points with the outer button on the arm, and confirm with the inner button. You can press the outer button several times until you are happy with the point and confirm with the inner button. You can move the arm after having pressed the outer button - this will not be registered.
- If an angular sensor reaches the end of its range, the arm will make a warning signal. You may then have to twist the arm around in some way to place the sensor into its working range.
- You can also correct mistakes later, by asking the program to step back in the list of points.
- When you are happy with the landmarks on one specimen, you can either save these in a separate file or continue with the next specimen.
- After loading the data file into PAST, you must first delete the columns with angular data etc., and then use the "Grouped rows to multivar" option in the "Edit" menu to put the numbers in the right order such that there is one specimen per row. PAST will ask for the number of rows per group - this will be the number of landmarks per specimen. You may also need to run "Commatize" from the "Edit" menu in order to convert decimal points to commas. Check the results using the "Plot landmarks (3D)" option.

We also have a much more advanced and complicated program, called "CAM Measure". One of the useful options in this program is the ability to digitize points each time the probe crosses a set of predefined, equidistant, parallel planes. This allows the acquisition of a complete 3D model of a specimen by "brushing" over its surface. For this application it is recommended to use a ball probe rather than the needle probe. Changing probe requires a recalibration of the arm. This is a simple procedure which is described in the manual.

One of the practical problems is the possible movement of a specimen during measurement. It is necessary to fix the specimen securely, for example using a vice. Also be aware of possible damage to the specimen from the metal needle - try not to drop the needle heavily onto the surface.

Coordinates from distances

Another interesting approach to acquiring 2D or 3D coordinates is to measure all distances between all landmarks with calipers. From this information, the absolute coordinates can be computed using Principal Coordinates Analysis as described in a later chapter, apart from a possible reflection (mirror image) which usually does not matter. One advantage of this method is that it does not require expensive equipment for 3D measurement. Also, the distances can be used directly by distance-based methods such as EDMA, which are presently competing with more traditional landmark methods. We will not discuss distance-based methods further in this course however.

The main disadvantage of this kind of coordinate acquisition is the large number of measurements needed if you have more than a few landmarks. For n landmarks there are $(n^2 - n)/2$ distances, so for ten landmarks there are 45 distances to be measured per specimen.

Accuracy

Measurement errors should be assessed by making multiple measurements, for example ten, ideally on every landmark on every specimen. The standard errors can then be computed, and the mean values are likely to be very accurate. If this is impractical, multiple measurements should at least be made on one specimen in order to get an idea about expected errors.

Chapter 3

Univariate hypothesis testing

If we have measured a single variable such as length or a length/width ratio on a number of specimens, we have available to us a range of classical, univariate statistical tests. We may want to test whether the measurements from two groups of samples (for example from two localities, stratigraphical positions or putative species) are significantly different, meaning that the difference is larger than what we might reasonably expect from inadequate sampling.

The procedure for carrying out such tests is usually to set up a null hypothesis of equality, which we may be able to reject and hence show inequality. For example, to test whether the mean length of *Tyrannosaurus* skulls is different from the mean length of *Tarbosaurus* skulls, we would set up the null hypothesis that the means are the same. Using an appropriate statistical test, we may then be able to show that the probability of equality of the means is less than 5 percent, which means that if you make two new groups by picking the same numbers of specimens as you have in your data set from the 'real', combined population (both genera) back in the Cretaceous, there is less than five percent chance of getting as large difference between the means as you have observed.

Already at this stage, I would like to state a few important points about such hypothesis testing:

- *Necessity of a statistical test.* Why not just compare the means without further complication? If the mean length of the *Tyrannosaurus* skulls is 1.23 meters and the mean length of the *Tarbosaurus* 1.02 meters, why can't we just state that the former was larger than the latter? We can't, because we haven't checked whether this difference could be the result of inadequate sampling. If you have a limited number of specimens, the means between the two groups will of course never be exactly the same, even if they come from the same population. To get an idea about whether the difference is larger than what could be reasonably expected from such sampling effects, you need to look at the variation within each group and compare with the difference between the means of the two groups. This is why we need a formal statistical test.
- *The null hypothesis can only be rejected, not confirmed.* Unfortunately, the logic and mathematics behind the statistical tests dictate that the null hypothesis can only be *rejected* (with a certain probability value), not confirmed. So if you test the null hypothesis of equality of the means, and get a *low* probability (say $p < 0.05$), you have a good result - you can reject the null hypothesis and have shown inequality. But if you get a high probability (say $p > 0.2$), you can only say that you are unable to reject the null hypothesis, and that you have no reason to claim inequality. In

spite of this formal state of affairs, a high p value will of course informally indicate equality. Just be careful to use the right 'code' when writing up your results for publication: say 'equality of the means could not be rejected ($p = 0.72$)' instead of 'the means were equal at $p = 0.72$ '.

- *Significance level.* At what level of probability should we set the threshold for a statistically significant result? This is to some extent a matter of personal taste. 0.01 is perhaps a bit too restrictive, and 0.1 a bit too sloppy, so 0.05 might be a reasonable compromise. In principle, one should choose a significance level before performing the test - it's considered bad practice to first get a probability level from the test and then choose a significance level which will give a result in accordance with your theory (but who will ever know?).
- *Circularity.* Your two (or more) groups to be compared must have been chosen from independent evidence. For example, let us say that you have a sample of ammonites and want to test for sexual dimorphism. It is no good to divide the sample into 'small' and 'big' specimens and then test statistically whether these two groups have different sizes! Of course they do - that's how the groups were selected in the first place. This basic error of circular reasoning is sometimes made by taxonomists wanting to test whether morphometric measurements indicate the splitting or merging of species.
- *Assumptions.* All statistical tests make a number of assumptions, and it is necessary to check whether it is reasonable to say that the assumptions hold for your data. In general, it is better to use the test which has the maximal number of assumptions, such as parametric tests which depend on a particular statistical distribution (usually the normal distribution). If you cannot meet the requirements of this test, you have to step down the ladder and look at tests with fewer assumptions, such as non-parametric tests. These tests are fine, but they often have less *power*, meaning that they may report a higher probability value than 'necessary', making it impossible to reject the null hypothesis even if there might be grounds for doing so.

Testing for normal distribution

Many statistical tests assume that the data are normally distributed (follow the 'bell' or 'Gauss' curve). Before carrying out such tests, it may be necessary to test for normality, that is, make sure that there is no reason to reject normal distribution. Whether the points are normally distributed can also be of inherent interest.

The normal distribution has two parameters: The mean and the variance. The variance indicates the width of the bell curve. The standard deviation is the square root of variance. About 68 percent of the data points will be within one standard deviation from the mean, while about 95 percent will be within two standard deviations.

There are several available tests for normal distribution. In PAST, you will find a good test with high power called Shapiro-Wilk, which is optimized for small sample sizes ($N < 50$). For larger samples you can use a Chi-squared test with somewhat poorer statistical power.

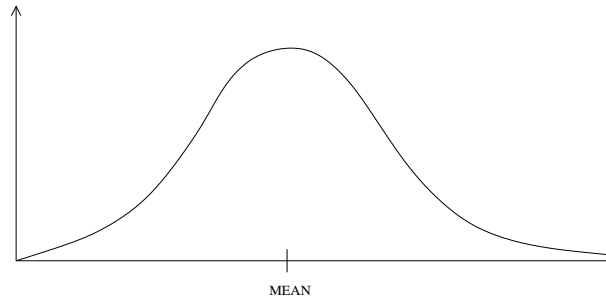


Figure 3.1: Normal distribution, also known as the 'Gauss' or 'bell' curve.

The F and the t tests

The F and t tests are the classics of statistics, and are very commonly used. These tests assume normal distributions. The F test is for equal variance, while the t test is for equality of the means. Note that the classical version of the t test assumes equality of variance. A slight modification known as the *Welch test* can be used in the case of unequal variances.

Mann-Whitney U, testing for equal medians

If we can not assume normal distributions, we can use a non-parametric test which is designed to handle any distribution. The most common one is the Mann-Whitney U, testing for equal medians. The median is the value such that there are as many data points with higher as with lower values. It will be equal to the mean value for symmetric distributions.

Kolmogorov-Smirnov, testing for similar distributions

The Kolmogorov-Smirnov is a non-parametric (distribution-independent) test for equality of distribution. It will indicate the probability that two samples have been taken from the same parent population with any distribution.

Permutation tests

An interesting alternative to the classical statistical tests is provided by randomization tests such as the permutation test for equality of the mean. In this test, the difference between the means of the two groups is first calculated. Then, a large number (often 1000) of random pairs of groups are produced, each by picking data points from the total (pooled) data set. If the difference between these groups is never as large as the difference between the original two groups, we can assume that the original groups are significantly different. By counting the number of times the randomized difference exceeds original difference, we can calculate a probability value for equality.

Such tests are more or less distribution-independent, have high power (do not over-estimate probabilities) and can be used for small samples.

ANOVA

ANOVA (analysis of variance) is a parametric test, assuming normal distribution, to investigate equality of the means of many groups. If you have measured sizes of ostracods, with many specimens from each of many localities, is the mean size the same at all localities? It turns out to be an incorrect statistical procedure to test this separately for all pairs of samples using for example the t test. The reason for this is simple. Say that you have 20 samples which have all been taken from a population with the same mean. If you try to use the t test for each of the 190 pairwise comparisons, you will most probably find several pairs where there is a 'significant' difference at $p < 0.05$, because such an erroneous conclusion is indeed expected in five percent of the cases at this significance level. ANOVA is the appropriate method in this case, because it makes only one total test for equality.

However, once overall inequality has been shown using ANOVA, you are allowed to proceed with so-called "post-hoc" tests in order to identify which pairs of samples are significantly different.

Traditional ANOVA assumes normal distributions and also equality of variances, but these assumptions are not critical as long as the samples are of the same size. If the sample sizes are different, the variation in variance should not exceed fifty percent. If any of these assumptions are violated, a non-parametric test known as the *Kruskal-Wallis test* is recommended.

Case study from PAST: Permian brachiopods

The Permian brachiopod *Dielasma* is common at many localities within the Zechstein reef facies of the Ford Formation in the Sunderland area, N.E. England (Hollingworth & Pettigrew 1988). The smooth terebratulide brachiopod often forms monospecific clusters which are presumed to represent living assemblages at the base and within the reef. Large, well-preserved samples of the genus are available and a number of variates may be easily measured on complete specimens. Hollingworth & Pettigrew (1988) assessed the population structures and dynamics of a number of discrete populations of *Dielasma* in terms of salinity and other possible environmental fluctuations. Study of several samples suggest some populations were stunted; moreover it is possible that several subspecies or even species may be represented in the samples associated with the Permian reef.

Two measurements were made, using vernier calipers, on three discrete samples of the terebratulide brachiopod *Dielasma* from the Ford Formation: the sagittal length (X1) and maximum width (X2) of the pedicle valve were taken on conjoined pairs. Data from three different sites (L1, L2 and L3) are available for analysis. With the information available it should be possible to compare and contrast the population structures and dynamics of each sample and moreover compare the relative outlines of the shells from all three localities.

We first want to investigate whether the X1 and/or X2 measurements are statistically different in the three samples. First, we should check whether the measurements are normally distributed. Since the samples are relatively large ($N > 30$ for all of them), we use the chi-squared test in PAST. For samples L1 and L2, the test reports $p < 0.05$. We can therefore reject normal distributions for these samples. For L3, we can not reject normality.

Since we have non-normality for two of the three samples, we can not use the F or t tests. We therefore choose the Mann-Whitney U test, which reports $p < 0.05$ for equality of the medians of any of the pairs (L1/L2, L1/L3, L2/L3). We can therefore be quite confident in assuming that the samples come

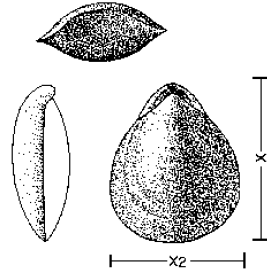


Figure 3.2: Measurements taken on *Dielasma*.

from different statistical populations.

It might be argued that this statistical procedure is slightly inappropriate, because we are comparing more than two groups. Some ANOVA-like tests do not require normality, and we should perhaps try one of those instead, but this concern is not serious for only three groups.

We will come back to this example in later chapters.

Chapter 4

Bivariate measurements, curve fitting and allometry

Curve fitting

Many data sets consist of pairs of measurements. Examples are lengths and thicknesses of a number of bones, grain sizes at a number of given levels in a section, and the number of species at different points in geological time. Such data sets can be plotted with points in a coordinate system (scatter plot). Often we wish to see if we can fit the points to a mathematical function such as a straight line or an exponential function, perhaps because we have a theory about an underlying mechanism which is expected to bring the observations into conformation with such a function. Most curve fitting methods are based on *least squares*, meaning that the computer finds the parameters that give the smallest possible sum of squared error between the curve and the data points.

Fitting to a straight line

The most common type of curve fitting consists in finding the parameters a and b that give the best possible fit to a straight line:

$$y = ax + b.$$

There are two forms of linear curve fitting. The most common type is *regression*, which assumes that the given x values are exact and independent from y , such that the measurement error or random deviation is only found in the y values. In the example of grain sizes we can perhaps assume that the level in meters is a given, almost exact, strictly increasing value, while the grain size is a more randomly varying variable.

The other form of linear curve fitting, called RMA (Reduced Major Axis) is to be preferred in the example of lengths and thicknesses of bones. The x and y values are here of more comparable nature, and errors in x and y are both contributing to the total squared error.

Regression and RMA can often give quite similar results, but in some cases the difference may be substantial.

Correlation, significance, error estimates

A linear regression or an RMA analysis will produce some numbers indicating the degree of fit. It should be noted that the significance value p and the estimation of standard errors on the parameters depend upon several assumptions, including normal distribution of the residual (distances from data points to the fitted line) and independence of the residual upon the independent variable. Least-squares curve fitting as such is perfectly valid even if these assumptions do not hold, but significance values can then not be trusted.

PAST produces the following values:

- $p(\text{uncorr})$: The probability that x and y are uncorrelated. If $p(\text{uncorr})$ is small ($p < 0.1$), you can use the values below.
- r : Correlation coefficient. This value shows the strength of the correlation. When x and y are increasing together, and are placed perfectly on a straight line, we have $r = 1$.
- $err a$: Standard error in a
- $err b$: Standard error in b

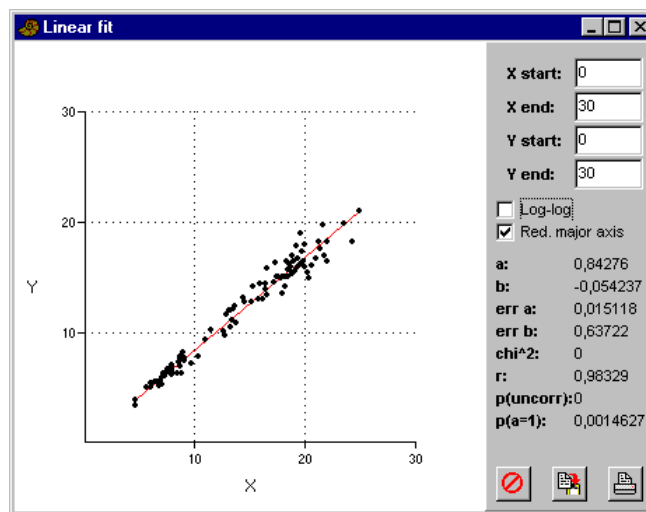


Figure 4.1: Example of linear regression. Note that in this case, the assumption of independence of the standard deviation of the residual upon the independent variable does not seem to hold well (the points scatter more for larger x).

Log and log-log transformation

We can use linear regression also for fitting the data points to an *exponential* curve. This is done simply by fitting a straight line to the logarithms of the y values (taking the logarithm transforms an exponential function to a straight line). If we use the natural logarithm, the parameters a and b from the regression are to be interpreted as follows:

$$y = e^b e^{ax}.$$

In PAST there is also a function for taking the base-10 logarithms of both the x and the y values. The data points are then fitted to the power function

$$y = 10^b x^a.$$

For the special case $a = 1$ we are back to a linear function.

Nonlinear curve fitting

All the functions above are linear in the parameters, or they can be linearized. It is more tricky for the computer to fit data to nonlinear functions. One example is the *logistic curve*

$$y = \frac{a}{1 + be^{-cx}}$$

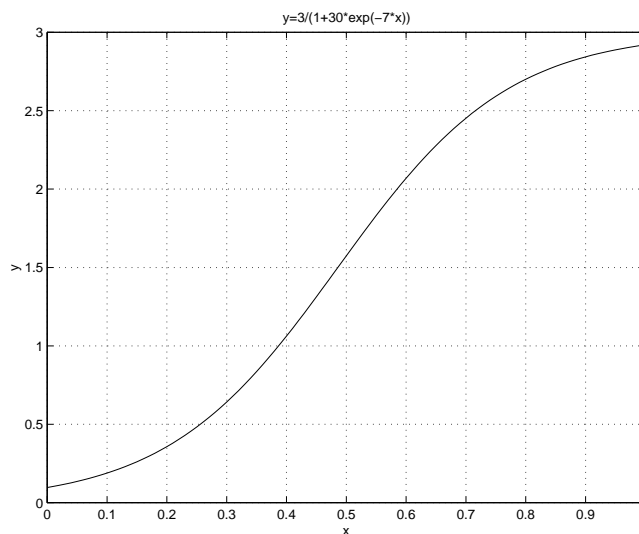


Figure 4.2: Logistic curve. $a=3$, $b=30$, $c=7$.

The logistic curve is often used to describe growth with saturation (fig. 4.2).

The question of whether, for example, a logistic curve fits a data set better than a straight line does is difficult to answer. We can always produce better fits by using models with more parameters. If Mr. A has a theory, and Mrs. B also has a theory but with more parameters, who shall we believe? There are formal ways of attacking this problem, but they will not be described here.

Allometry

When taking two linear measurements such as length and width on a number of individuals at different growth stages, it will sometimes be found that the ratio between them remains constant such that the shape stays the same. This is called *isometric* growth, meaning that growth rate is the same everywhere (homogenous) and the same in all directions (isotropic). But it is more commonly seen that the ratio between the measurements changes systematically through ontogeny, giving a continuous change in proportion. This is known as *allometric* growth.

A well-known example of allometric growth is head size in humans and many other vertebrates. The head grows slower than the rest of the body, giving a diminishing ratio between head and body size as the child matures. This fact is used to advantage by the toy industry, invoking an emotional response by making the heads of teddy bears and other figures disproportionately large (fig. 4.3).



Figure 4.3: Plastic ducks purchased in Oslo, showing distinct allometry. The juvenile has a large head relative to body (ratio 0.48 compared with 0.42 in the mother, further emphasized by the head being placed further back), and a small bill relative to head (ratio 0.27 compared with 0.44 in the mother),

Imagine an evolutionary process where the timing of ontogeny is being altered. If the growth rates of head and body become more similar, it will take longer time for the head to become small relative to the rest of the body through ontogeny. Thus, small head size is delayed to progressively higher age of the individual. This is what happened in the evolution from non-human primate to man, causing larger head size in the adult. Such change in developmental timing is called *heterochrony*, and it is an exceedingly important evolutionary process (although not the only one, as some have almost claimed!).

The phenomenon of heterochrony as a crucial phylogenetic process makes it even more interesting and necessary to identify allometry in detail. This can only be done using morphometry.

Not only will the ratio between two measurements x and y change through ontogeny, but it will normally do so in a very specific way, such that the *logarithms* of the measurements stay in constant proportion. Therefore we will normally see that if we plot the bivariate measurements in a diagram with both axes being logarithmic, the points will end up almost on a straight line. As mentioned above, we can therefore take the logarithms of all our measurements and perform linear regression or preferably

RMA analysis, fitting the points to the *allometric equation*

$$y = bx^a.$$

If $a = 1$, we are back to a linear relationship between x and y , and we have isometric rather than allometric growth.

In the next section, which can be skipped on first reading, I explain why this seemingly strange result makes perfect biological sense in terms of relative growth rates. Growth rate should preferably be measured relative to size, because growth happens by cell division such that a constant division rate will give faster overall absolute growth in a larger body. If we view growth rates in this way, it is easy to see that non-homogenous growth will follow the allometric equation.

Derivation of the allometric equation

We have two linear distances x and y , with growth rates dx/dt and dy/dt . We then assume that the relative growth rate of y is a constant a multiplied with the relative growth rate of x (y grows faster than x if $a > 1$):

$$\frac{dy/dt}{y} = a \frac{dx/dt}{x}. \quad (4.1)$$

We cancel dt and get

$$\frac{1}{y}dy = a \frac{1}{x}dx. \quad (4.2)$$

Integration on both sides gives

$$\ln y = a \ln x + C \quad (4.3)$$

for an arbitrary constant of integration C , meaning that the logarithms of the two measures plot on a straight line. Taking the exponential on both sides, we get

$$y = e^C x^a.$$

We let e^C change name to b , and finally get the *allometric equation*

$$y = bx^a.$$

The exponent a is then the ratio between the relative growth rates of x and y . For $a = 1$ we have isometric growth.

a can often change through ontogeny, and if this happens stepwise the plot of $\log x$ against $\log y$ will be piecewise linear.

Log-transformation of morphometric data

Many data analysis methods, including PCA and discriminant analysis which will be described in later chapters, work best if the variables have some degree of linear relationship to each other. However, in the case of allometric growth and heterochrony, we have seen that distances will rather be related through an allometric equation. Some authors therefore suggest and recommend that all distance measurements are converted by taking their logarithms prior to analysis. Such *log-transformation* will linearize the relationships between variables. Other workers feel that log-transformation makes the analysis more indirect and the results more complicated to interpret.

Permian brachiopods revisited

We can use the *Dielasma* data set described earlier to investigate possible allometry. We log-transform the X1 and X2 values, and perform an RMA analysis on the result. The null hypothesis of isometric growth ($a = 1$) can not be rejected at a $p < 0.05$ significance level for any of the three samples, so we must conclude that we have no reason to claim allometry in this case.

Chapter 5

Asymmetry

The developmental mechanisms responsible for biomorphological symmetry are still only partly known. However, it is clear that perfect regulation of growth at each side of the body is necessary to ensure bilateral symmetry, and that symmetry therefore requires a high degree of developmental stability. Therefore, it has for a long time been assumed that the degree of asymmetry in the phenotype could indicate something about the amount of inherent (genetic) and extrinsic (environmental) stress during development. That symmetry could be a sensitive indicator of genetic and environmental quality, and that females prefer symmetrical males, has become one of the most famous ideas of modern biology.

Although this field has recently been criticized for producing some questionable results, the fact remains that asymmetry is an interesting morphological parameter which can potentially give important clues about ancient environments, developmental processes and genetic "events" such as hybridization. However, the degree of asymmetry is usually so small (perhaps 1-2 percent) that very careful measurements must be taken, and diagenetic and tectonic deformation can be a severe problem.

Measuring asymmetry

Asymmetry can be measured in two fundamentally different ways, by using either *meristic* or *metrical* traits (Palmer 1994).

Meristic traits are the number of discrete elements such as sensory bristles, fin rays, rows of scales, tubercles or ribs. The great advantage of using this kind of data is that the counts can be expected to be reasonably accurate, and independent from post-mortem deformation of the specimen. The main disadvantage is that such counts can be statistically useless if the number of elements normally differ by only one or two between the sides (see Palmer 1994).

Metrical traits are based on continuous, linear measurements.

Both meristic and metrical asymmetry is measured by calculating the difference $R - L$ in count or distance between the left and the right side. This will have to be done on a population of specimens, preferably more than 30, so that a histogram of frequency distribution can be drawn. Measuring asymmetry at one growth stage in a single individual is not a useful exercise.

Antisymmetry, directional asymmetry, fluctuating asymmetry

We can distinguish between three main types of asymmetry distribution in a population: Antisymmetry, directional asymmetry and fluctuating asymmetry. Combinations of these basic patterns are possible.

Antisymmetry means that most individuals are substantially asymmetric, but the "handedness" is random. The R-L distribution will be bimodal. For example, most lobsters have one claw much bigger than the other, but whether this is the left or the right claw is more or less arbitrary, and in fact determined by the environment.

Directional asymmetry means that most individuals are asymmetric, and to one particular side. The sizes of the hemispheres of human brains is a typical example - most people have a larger left hemisphere. The R-L values follow a normal distribution with a non-zero mean.

Fluctuating asymmetry is perhaps the most common case, and the one most easy to study with statistical methods. In this case, the "standard" configuration is symmetric, and the R-L values follow a normal distribution with a zero mean. The standard deviation (width of the bell curve) then indicates the degree of asymmetry in the population, that is, the extent of deviation from the "ideal" morphology. In this case, the standard deviation of R-L has been argued to be a good indicator of developmental stability.

Palmer (1994) gives a thorough, practical introduction to statistical procedures for the study of fluctuating and directional asymmetry.

Chapter 6

Cluster analysis

Cluster analysis means finding *groupings* of specimens, based on an appropriate distance measure. Such groups can then be interpreted in terms of environment, age or taxonomy. *Hierarchical* cluster analysis will produce a so-called *dendrogram*, where similar specimens are grouped together. Similar groups are further combined in 'superclusters', etc. (fig. 6.1).

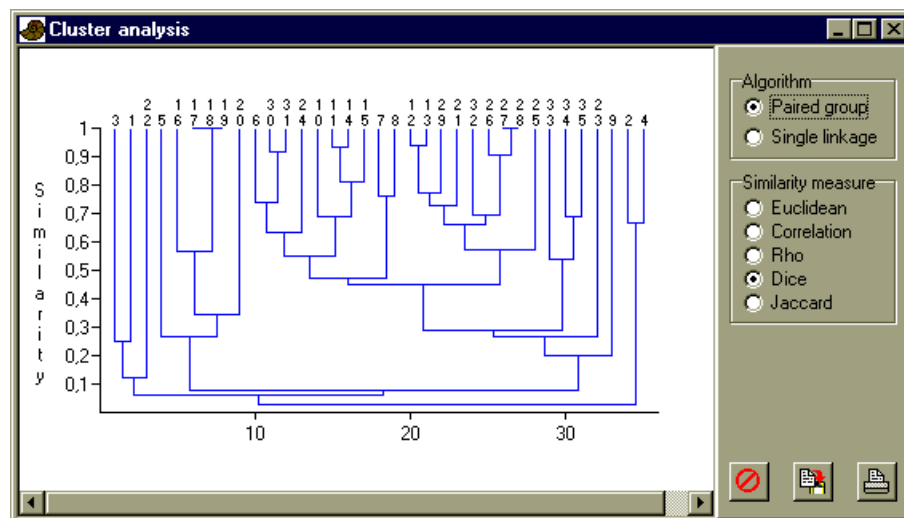


Figure 6.1: Dendrogram. The vertical axis is in units of group similarity.

Using cluster analysis of specimens, we can for example see whether they can be grouped into morphotypes which may again be interpreted in terms of environment or taxonomy.

There are several algorithms available for hierarchical clustering. Most of these algorithms are *agglomerative*, meaning that they cluster the most similar items first, and then proceed by grouping the most similar clusters until we are left with a single, connected supercluster. In PAST, the following algorithms are implemented:

- Mean linkage, also known as Unweighted Pair-Group Moving Average (UPGMA). Clusters are joined based on the average distance between all members in the two groups.

- Single linkage (nearest neighbour). Clusters are joined based on the smallest distance between the two groups.
- Ward's method. Clusters are joined such that increase in within-group variance is minimized.

For morphometric analysis, the UPGMA algorithm or Ward's method is recommended. Ward's method seems to perform better when the Euclidean distance measure is chosen. It may however be useful to compare the dendrograms given by the different algorithms and different distance measures in order to informally assess the robustness of the groupings. If a grouping is changed when trying another algorithm, that grouping should perhaps not be trusted.

Morphological cluster analysis and taxonomy

In olden days, cluster analysis based on morphological characters or measurements was used in systematics at supraspecific levels. This was called 'numerical taxonomy'. With the rise of cladistics, this approach was seen as theoretically flawed for a number of reasons, the most important one being that it was meaningless in terms of phylogeny.

When we use cluster analysis and other multivariate methods on morphological data these days, it is therefore normally not with the aim of producing a supraspecific taxonomy, but to study *intraspecific variation*, or possibly to investigate microevolution or distinguish between morphospecies. For these purposes, cluster analysis can be a powerful tool. It allows us to discover more or less separated morphological groups, which may be interpreted in terms of environment, genetic differences, sexual dimorphisms, instars or year classes.

Chapter 7

Principal components analysis

Principal components analysis (PCA) is a method that produces hypothetical variables or *components*, accounting for as much of the variation in the data as possible (Jolliffe 1986). The components are linear combinations of the original variables. This is a method of data reduction that in well-behaved cases makes it possible to present the most important aspects of a multivariate data set in two dimensions, in a coordinate system with axes that correspond to the two most important (principal) components. In other words, PCA represents a way of projecting points from the original, high-dimensional variable space onto a two-dimensional plane, with a minimal loss of information.

In addition, these principal components may be interpreted as reflecting underlying morphological variables with a biological meaning (fig. 7.1).

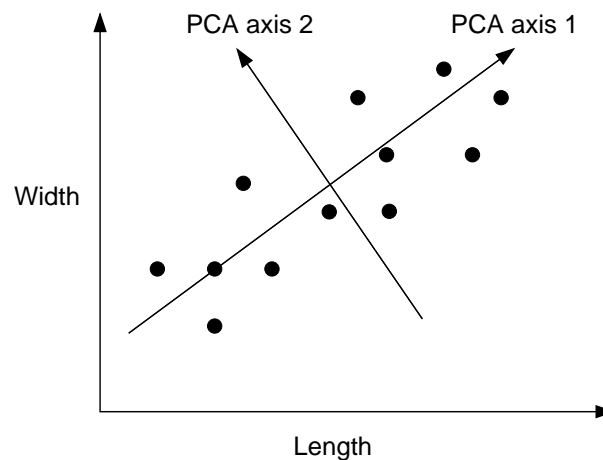


Figure 7.1: Hypothetical example of PCA. The lengths and widths of 12 dinosaurs have been measured. The 12 specimens are shown in a scatterplot as usual. PCA implies constructing a new coordinate system with the sample centroid at the origin and with axes normal to each other such that the first axis explains as much of the variation in the data as possible. In this case, we would interpret axis 1 as size, and axis 2 as a width/length ratio.

PCA is tricky to grasp in the beginning. What is the meaning of those abstract components? Consider

the following example. We have measured shell size x , shell thickness y and a colour index z on 1000 foraminiferans of the same species but from different climatic zones. From these three variables the PCA analysis produces three components. We are told that the first of these (component A) can explain 73 percent of the variation in the data, the other (B) explains 24 percent, while the last (C) explains 3 percent. We then assume that component A represents an important hypothetical variable which may be related to environment.

The program also presents the 'loadings' of component A, that is how much each original variable contributes to the component:

$$A = -3.7x + 1.4y + 0.021z.$$

This tells us that A is a hypothetical variable that reduces sharply as x (shell size) increases, but increases when y (shell thickness) increases. The colour index z has almost no correlation with A. We guess that A is an indicator of *temperature*. When temperature increases, shell size diminishes (organisms are often larger in colder water), but shell thickness increases (it is easier to precipitate carbonate in warm water). Plotting the individual specimens in a coordinate system spanned by the first two components supports this interpretation: We find specimens collected in cold water far to the left in the diagram (small A), while specimens from warm water are found to the right (large A).

It is sometimes argued that PCA assumes some statistical properties of the data set such as multivariate normality and uncorrelated samples. While it is true that violation of these assumptions may degrade the explanatory strength of the axes, this is not a major worry. PCA, like other indirect ordination methods, is a descriptive method without statistical significance anyway. There is no law against making any linear combination of your variables you want, no matter the statistical properties of your data, if it is found to be useful in terms of data reduction or interpretation.

Case study from PAST: Illaenid trilobites

Although distinctive and often common in Ordovician faunas, illaenid trilobites have relatively featureless, smooth exoskeletons, and taxonomic discrimination within the group is therefore difficult. While a number of authors have attempted a qualitative approach to illaenid systematics, Bruton & Owen (1988) described the Norwegian Upper Ordovician members of the family in terms of measurements and statistics. We will use multivariate methods to try to discriminate between at least two putative species of the illaenid trilobite *Stenopareia*.

As shown in figure 7.3, four simple measurements (L2, W1, W2 and W3) were taken on *Stenopareia glaber* from the Ashgill of Norway together with *S. linnarssoni* from both Norway and Sweden. A previous study of *S. glaber* recognized 'long' and 'short' forms.

We include all the specimens in one PCA analysis, identifying major axes of morphological variation. The first principal component is overwhelmingly important, explaining 98.8 percent of the variation. The loadings (weights on the four variables) for the first principal component are shown in figure 7.4. We see that as we go along this axis, all four distances increase almost equally fast. This must mean that the first axis captures general *size*. In fact, this will happen for almost all morphometric data sets, unless normalized for size. Size will practically always be responsible for most of the morphometric variation, and will therefore turn up as the first principal component. Some authors even *define* size as the score on



Figure 7.2: Norwegian illaenid trilobite. Paleontological Museum, Oslo. Length approx. 6 cm.

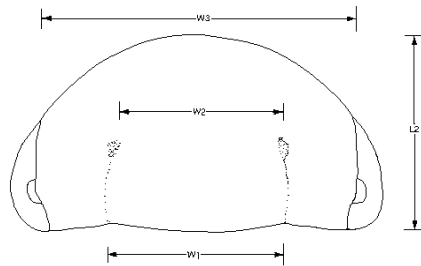


Figure 7.3: Four measurements made on the crania of the Norwegian illaenid trilobites are defined on the figure.

the first component of PCA. To study shape, we must go beyond the uniform scaling represented by the size axis.

The loadings on the second component show that as we go along the second axis, length is rapidly decreasing while the width values are increasing. The second component can therefore be described as a length/width ratio axis. Using only the first two components, we have reduced the original four-dimensional data set to two dimensions. PCA identifies the best way of performing this reduction, losing a minimal amount of information. We can now plot all the specimens in principal coordinate space, with the first component (length) along the horizontal axis and the second component (length/width) along the vertical (figure 7.5). Plotting length against length/width would normally be considered somewhat silly, but we do it in order to clarify the separation into two components with very different numerical importances.

The scatter plot shows that the three groups to some extent occupy different regions of morphospace, which is interesting, but there is no clear separation between them. This makes it difficult to use these measurements alone for erecting species.

Principal coordinates analysis

Principal coordinates (PCO) analysis starts from distance values between all pairs of specimens using any distance (or similarity) measure. The points are then placed in a low-dimensional space such that the distances are preserved as far as possible. PCO is also known as metric multidimensional scaling

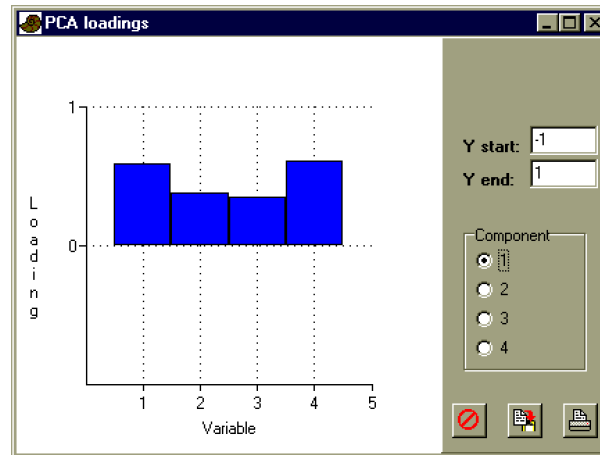


Figure 7.4: Loadings on first principal component.

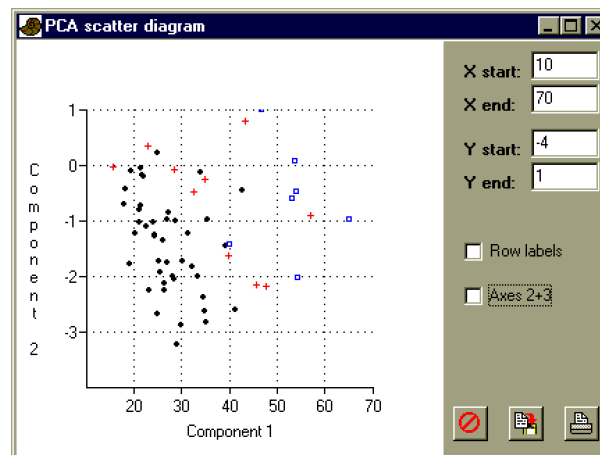


Figure 7.5: Scatter plot of the three groups in principal coordinates space.

(MDS).

PCO is attractive because it allows the use of any distance measure, not only the Euclidean.

Non-parametric multidimensional scaling

Non-parametric multidimensional scaling (NMDS, Kruskal 1964) starts from a ranking (ordering) of distance values between all pairs of data points using any distance measure. These ranks are then used in an iterative procedure in order to try to place the points in a low-dimensional space such that ranked distances between all pairs of points are preserved. One technical problem is that available algorithms do not guarantee an optimal ordination solution.

Since NMDS intentionally disregards absolute distances and only uses their ranks, it is ideal for situations where the distance measure is poorly understood. It is therefore popular in fields such as

ecology and economy, but may be less appropriate for many problems in morphometry where we have well-defined distance measures.

Chapter 8

Multivariate hypothesis testing and discriminant analysis

Multivariate normal distributions and confidence ellipses

The normal distribution can easily be extended to the case when we have more than one variable. The multivariate normal distribution is characterized by a multivariate mean or *centroid*, a separate variance in each of the variables, and also the *covariances* between the variables. For the bivariate case, the "bell" then turns into a function of two variables - a "hat", with a certain position (mean), certain widths along the two axes (variances) and a certain rotation depending upon the covariance.

Such a bivariate normal distribution will have *confidence ellipses* at different p values. For example, 95 percent of the data points will be found within the 95 percent confidence ellipse. Assuming bivariate normal distribution, the confidence ellipse can be calculated for any two-dimensional sample. It can be very useful to draw such ellipses onto a scatter diagram of the data, showing at a glance the variances, the mean (center of the ellipse), and the covariance (orientation of the ellipse). The confidence ellipses are particularly useful in scattergrams from PCA and other ordination methods.

Multivariate hypothesis testing

There are many statistical tests available for equality between multivariate samples. Hotelling's T^2 for equality of multivariate means is quite popular, being an analogue to the univariate t test. It assumes multivariate normal distribution.

Just like Hotelling's T^2 can be viewed as the multivariate extension of the t test, there is also a multivariate version of ANOVA, appropriately named MANOVA (Multivariate ANalysis Of VAriance). MANOVA should be used when we have several groups, each with multivariate normal distribution, and we want to test whether they all have the same multivariate mean. In the same way as ANOVA assumes similar variances in all the groups, MANOVA assumes similar variances and covariances. This can be investigated with another multivariate test known as Box's M. There are several test statistics in use as part of MANOVA. *Wilk's lambda* is perhaps the most common one, but the *Pillai trace* is also recommended.

It could well be that randomization (permutation) tests should be used rather than Hotelling's T^2 and MANOVA, since they make basically no assumptions. This is however not yet commonly seen in publications.

Discriminant analysis and Canonical Variates

If we have multivariate morphometric measurements from two groups of individuals, we may want to investigate whether the groups can be well separated or blend into each other in a continuum. One method for doing this in an informal manner is *discriminant analysis* (DA). DA produces an axis through the data, along which the separation is maximal. It can be compared with PCA, and the discriminant axis can be similar to the first PCA axis if separation is large. But in general it is a different type of method, maximizing separation between two *à priori* known groups instead of maximizing pooled variance. Discriminant analysis can also be extended to more than one axis, so that specimens can be plotted in scatter plots of discriminant scores in the same way as PCA scores. This is known under different names having the word 'canonical' in it, such as Canonical Variates Analysis (CVA). This type of analysis can also be extended to more than two groups.

CVA is a good method of data reduction when all specimens can be unequivocally assigned to one group. However, if group membership is uncertain or based on the variables used in the analysis, it might be better to use PCA in order to produce an ordination using fewer assumptions.

Example: Illaenid trilobites revisited

Returning to our trilobite example, we can plot all our specimens in a histogram along the discriminant axis (figure 8.1). The two groups are not totally separated, meaning that we may well find specimens which cannot be unequivocally placed in either group from these measurements alone. We may therefore conclude that these measurements provide insufficient information for the splitting into two species (there are however other characters which support such a split).

While discriminant analysis shows unsatisfactory *separation* of the two groups, the Hotelling's T^2 test indicates a highly significant *difference* between them.

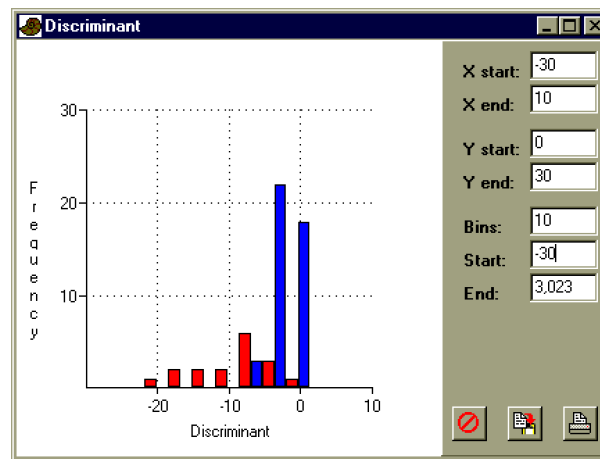


Figure 8.1: Histogram of the illaenid trilobites along the discriminant axis. *Stenopareia linnarssoni* red, *S. glaber* blue.

Chapter 9

Methods for some special types of morphometric data

Directional analysis

Measurements of orientations of fossils or sedimentary structures are important in sedimentology, being used to make inferences about paleocurrents and wave action. Orientations are less commonly used in biological morphometry, but have applications in the study of ornamentation (e.g. terrace lines in arthropods), osteological structures in vertebrates, and trace fossils. The statistical treatment of directions or orientations is slightly different from that of measured linear variables, because of the periodic nature of angles: 361 degrees is closer to 0 degrees than to 180 degrees. Therefore, special statistical tests must be used.

Rayleigh's test (Davis 1986) is used for testing the null hypothesis of uniform random distribution versus an alternative hypothesis of one single preferred direction. A simple chi-squared test may also be used, testing the null hypothesis of uniform distribution (equal number of directions in each of a small number of equal-sized bins).

Angular measurements

Angular measurements, either made directly with a goniometer or computed from landmarks, are sometimes used in morphometry, although it is in many cases doubtful whether this offers any advantages over distance measurements. Since angles need slightly different statistical treatment than distances, it is recommended to convert angles into linear measurements, especially if a mix of angular and linear measurements are included in the data set. A simple trick is to take the *cosine* of the angles prior to analysis. This is equivalent to converting the angle into a ratio of linear measurements in a hypothetical right-angled triangle.

Point distributions

Are points randomly distributed, clustered or overdispersed? We will concentrate on the two-dimensional case, and set up the null hypothesis that the points are randomly, uniformly and independently distributed. We analyze the points by studying the distribution of the distance from each point to its *nearest neighbour*. The theoretical mean of nearest neighbour distances for a given number of points in a given area under the random hypothesis is known, and can be compared with the observed mean. The difference of the means can be tested statistically (Davis 1986), and a dispersion parameter R defined:

$$R = \frac{2\bar{d}}{\sqrt{A/N}},$$

where \bar{d} is the mean nearest neighbour distance, A is the area and N is the number of points. Clustered points (mean nearest neighbour distance smaller than expected from the null hypothesis) will give $R < 1$, while overdispersion, presumably due to lateral inhibition, will give $R > 1$.

The test assumes that a) the number of points is so large that 'boundary effects' are small ($N > 50$), b) the points are really points, that is, their diameters are very small compared with the distances between them, and c) the area can be accurately estimated, for example using the convex hull (see the section on size from landmarks).

These assumptions can be partly avoided by using a randomization test, generating a large number of random point patterns in a similar geometry and comparing their mean nearest neighbour distances to the observed one.

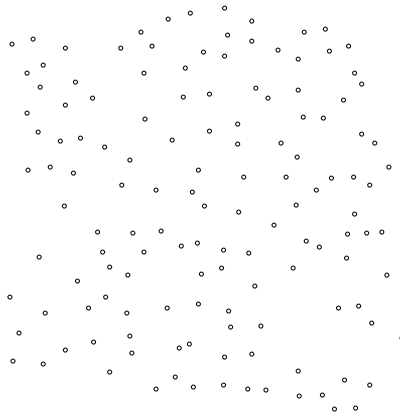


Figure 9.1: Cranial tubercles from the large Cambrian trilobite *Paradoxides forchhammeri*, Norway. $p(\text{random}) < 0.05$, $R = 1.28$. The overdispersion indicates a lateral inhibition mechanism, quite possibly through the action of the *Delta* and *Notch* gene products responsible for the spacing of sensory hairs in arthropods (Hammer 2000)

Chapter 10

Outline analysis

From analysis of one or a few distances, we will now proceed to methods for studying the variability of complete *shape*. How can complex shapes be quantified and reduced to a manageable data set?

In this chapter, we will look at outline or contour analysis, based on a digitized line (large number of points) around the edge of an object. There is an ongoing argument about whether such outlines are useful to analyze, and in particular whether outline analysis will be comparing homologous structural elements or not. Some authors dismiss outline analysis, and recommend landmark analysis instead (next chapter). As pointed out by MacLeod (1999) this discussion seems a little formalistic. In many cases we do not have good landmarks, and outline analysis may be the only option for studying shape anyway.

Fourier analysis

Several methods for outline analysis is based on Fourier¹ analysis, which must be briefly explained.

Consider any periodic function of one independent variable, that is a function that repeats itself at regular intervals. A cardiogram of a healthy heart, sunspot activity through the eleven-year sunspot cycles and air pressure as a function of time in the sound of a pure, musical tone are all examples of approximately periodic functions. The duration of one cycle is called the *period*, which we will denote by T , and its inverse is *frequency* ($f = 1/T$).

It can be shown that you can produce any such periodic function by adding together sine and cosine functions with appropriate multiplication factors (amplitudes). In other words, you can always make a linear combination of sines and cosines that will reproduce your periodic function. Moreover, the frequencies of these sines and cosines will be multiples of the frequency of the original function. So any periodic function $g(x)$ with a frequency f can be expressed as follows:

$$g(x) = a_0 \cos(0 \cdot 2\pi f x) + b_0 \sin(0 \cdot 2\pi f x) \tag{10.1}$$

$$+ a_1 \cos(1 \cdot 2\pi f x) + b_1 \sin(1 \cdot 2\pi f x) \tag{10.2}$$

$$+ a_2 \cos(2 \cdot 2\pi f x) + b_2 \sin(2 \cdot 2\pi f x) \tag{10.3}$$

$$+ \dots \tag{10.4}$$

¹Jean Baptiste Joseph Fourier, 1768-1830, developed this field as part of his work on heat transfer

$$= a_0 + \sum_{i=1}^{\infty} [a_i \cos(2\pi i f x) + b_i \sin(2\pi i f x)]. \quad (10.5)$$

Conversely, given a periodic function, we can use *Fourier analysis* to decompose it into its constituent sines and cosines (the harmonics), each at a multiple of the fundamental frequency f .

Fourier shape analysis

A simple, closed outline may be given in polar coordinates, with radius as a function of angle relative to a fixed internal point such as the centroid. It is obvious that this function is periodic, repeating every full revolution of the polar vector. Now how can we reduce such a periodic function to a few numbers, describing the most important aspects of the shape? A natural idea is to use Fourier analysis to decompose the periodic function into its simple harmonic components, which are in the form of a harmonic series of sines and cosines. It turns out that many shapes, unless they include a large number of sharp corners and spikes, can be adequately described by the amplitudes of only a few of the first (low frequency) harmonics.

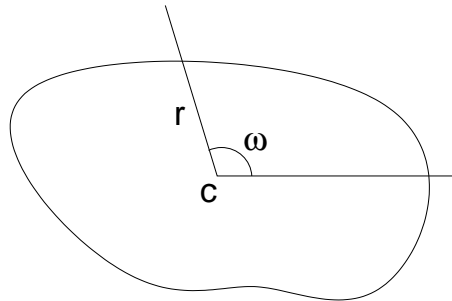


Figure 10.1: The principle of simple Fourier shape analysis. The closed outline is expressed as radius r as a function of angle ω . The centroid c is the origin of the coordinate system. The function $r(\omega)$ is periodic with a period of 2π , and can therefore be subjected to Fourier analysis.

Having reduced our data set consisting of a large number of digitized points to a smaller number of parameters describing the shapes, we can study the distribution of the shapes in a morphospace having the Fourier coefficients along the axes. However, since this space will still be quite high-dimensional (perhaps 20 dimensions), it will normally be necessary to reduce dimensionality further, for example using PCA.

Elliptic and tangent Fourier shape analysis

One serious problem with simple Fourier shape analysis is its use of a single-valued radius as a function of angle. In a shape of any complexity, there will not exist a fixed interior point from which we can draw a straight line out to any point on the outline without crossing it more than once. Simple Fourier analysis will fail for such shapes.

More advanced techniques for Fourier shape analysis eliminate this problem. First, we can simply let radius be a function not of angle, but of distance around the outline from a given starting point. Such perimeter-based Fourier shape analysis is simple and appealing, but has not been much used.

The most popular form of Fourier shape analysis is undoubtedly so-called Elliptic Fourier Analysis or EFA (Ferson et al. 1985). This method proceeds by interpolating the outline to get a large number of equally spaced points. Stepping around the outline from a given starting point, we take note of the x and y increments from point to point (fig. 10.2). These increments will define two periodic functions which are independently subjected to Fourier analysis. EFA seems to work well, but it has the disadvantage of producing twice as many parameters as other types of Fourier shape analysis. Such profusion of parameters seems unnecessary when considering that $x^2 + y^2$ will be constant for a given step size.

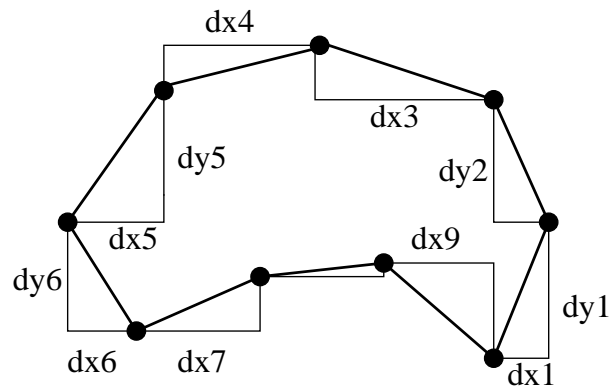


Figure 10.2: Elliptic Fourier analysis of a closed contour. The outline has been interpolated so that 9 points are placed at regular distances. The counterclockwise increments in the x and y directions define two periodic functions which can be subjected to Fourier analysis.

Again, the idea is that the original, high-dimensional shape space as defined by digitized points is reduced to a lower-dimensional space which retains the most important shape information (fig. 10.3).

An obvious variant of EFA is to take note of the increments not in Cartesian coordinates, but in polar coordinates. The angular directions of the increments, being a periodic function, is subjected to Fourier analysis, while the constant length of increments is discarded as a simple scaling factor. This is called Tangent Fourier Analysis.

Case study from PAST: Devonian trilobites

Crônier et al. (1998) studied the ontogeny of the Late Devonian phacopid trilobite *Trimerocephalus lelievrei* from Morocco. Well-separated size classes (growth stages) were identified, and the changing shape of the cephalon through ontogeny was studied using Elliptic Fourier shape analysis. This case study uses the original outline data set, kindly provided by Catherine Crônier.

As seen in figure 10.3, the outlines are well captured by the elliptic Fourier coefficients. MANOVA on the coefficients indicates that the shapes are not the same in the different growth stages (Wilk's $\lambda = 0.0015$, $p < 0.001$).

A PCA score scatter plot of the coefficients is shown in figure 10.4, where each specimen plots as

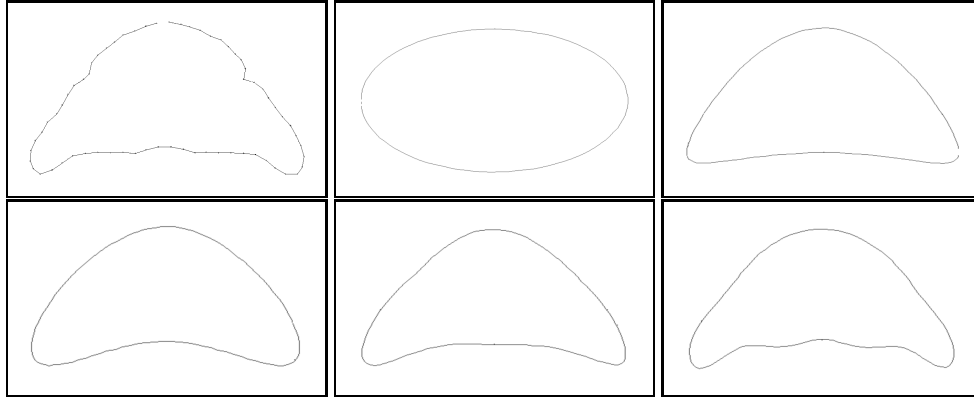


Figure 10.3: A trilobite cephalon. Original, digitized outline with 64 points (data from Cronier et al. 1998), and outline reconstructed with 1, 2, 3, 5 and 9 harmonics from Elliptic Fourier Analysis. Each harmonic is specified by four parameters (the sine and cosine factors for the x and y increments). The outline reconstructed with nine harmonics is therefore completely specified by $4 \times 9 = 36$ numbers, compared with the original $64 \times 2 = 128$ numbers.

a point in a shape space of reduced dimensionality. In this plot, we have also included the 95 percent confidence ellipses of each growth stage (the first growth stage is represented by a single specimen marked with a square). These ellipses are the bivariate versions of univariate confidence intervals. The overlap between the ellipses is large, but there is a definite trend from the upper left (first growth stage) to the lower right (last growth stage). The two last growth stages have almost identical positions, indicating that the morphology of the cephalon has stabilized at maturity.

Another interesting observation is that there seems to be a large jump between growth stage 3 and 4, corresponding to a change in molting mode. Also, the intraspecific variation seems to be smaller in stage 3 than in 2, and smaller in stage 5 than in 4, possibly indicating developmental stabilization (canalization) within each molting mode.

Eigenshape analysis

Eigenshapes² (MacLeod 1999) are the principal components of a set of shapes. The original shapes can then be given as linear combinations of the eigenshapes, and plotted as points in a PCA scatterplot. Also, the eigenshapes themselves can be drawn, and their variation under scaling will reveal the geometrical meaning of the PCA axes.

Eigenshape analysis can proceed in a number of ways, but the most common procedure goes as follows. First, the outline is resampled (interpolated) at equal distances. The turning angle for each increment is noted, much as in Tangent Fourier Analysis. Each shape is therefore given as a many-dimensional vector, and the vectors from all shapes are then subjected to PCA based on the covariance

²The name derives from the fact that eigenshapes are the eigenvectors of the shape covariances. Eigenvectors (and eigenvalues) are instrumental in many multivariate analysis methods, including PCA and CVA. "Eigen" comes, of course, from German, and isn't really meaningful in this connection, but still seems to make perfect sense. Attempts at translating the term into English are sometimes seen, but have not become popular

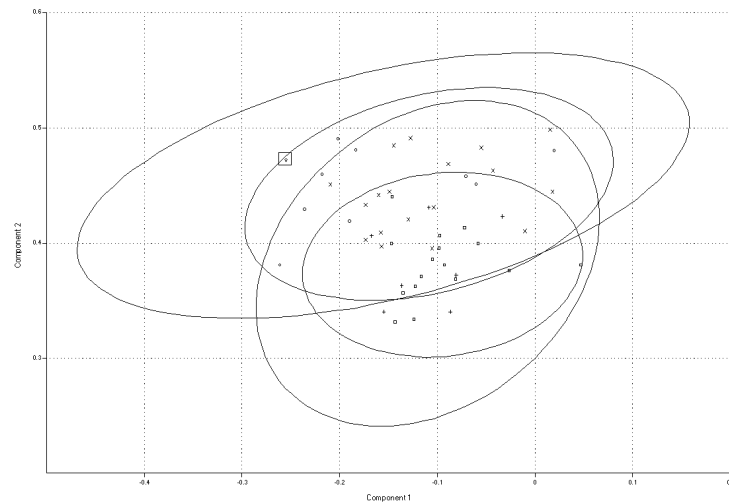


Figure 10.4: PCA scores for the first two principal components of the elliptic Fourier coefficients of trilobite cephalon data. A 95 percent confidence ellipse has been calculated for each group.

matrix. In contrast with Fourier methods, eigenshape analysis can be used also for open curves, such as the cross sections of bivalves.

Chapter 11

Landmark-based methods

Landmarks are well-defined points which are supposedly homologous from one specimen to the next¹. Good landmarks are typically placed at sharp corners, at the pointed ends of elongated structures, or at very small structures which may be regarded as points. Shape analysis based on landmarks has several advantages over outline-based methods:

- *Homology*. By measuring the same landmarks on all specimens, we are comparing like with like.
- *Ease of data collection*. It is normally less work to digitize a few landmarks than tracing a complete outline.
- *Ease of analysis*. There are many good methods available, in addition to simple analysis of distances between selected landmarks. The results of landmark analysis can sometimes (not always!) be relatively easy to interpret in terms of geometry and hopefully biology.

A thorough, but somewhat technical introduction to landmark-based morphometrics is given by Dryden & Mardia (1998).

Size

What is the size of a specimen? One may get different results depending on what measure is used - length (including the tail or antennae?), width, area or volume. A standard size measure based on landmarks is the *centroid size*, which is computed as follows. First, the centroid of all the landmarks is found. The centroid is the point having x and y coordinates equal to the mean of the x and y coordinates of all the landmarks. Then, the squared distances from the centroid to all landmarks are summed, and the square root taken. If we have a set of landmarks (x_i, y_i) with centroid (\bar{x}, \bar{y}) , the centroid size is given by

$$CS = \sqrt{\sum_i (\bar{x} - x_i)^2 + (\bar{y} - y_i)^2} \quad (11.1)$$

¹There has been some discussion about the issue of homology between landmarks. It could be argued that the concept of homology should be reserved for the phylogenetic equivalence of *structures* rather than mathematical points (MacLeod 1999)

An alternative size measure is the square root of the area of the *convex hull* of the points. The convex hull is the smallest possible convex polygon such that all landmarks are either at the vertices of the polygon or inside it (fig. 11.1).

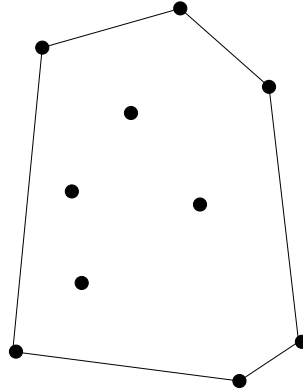


Figure 11.1: A set of landmarks, and its convex hull. The area of the convex hull (or its square root) can be used as a size measure.

Devonian trilobites revisited

Returning to the data set from Crônier et al. (1998) studied in a previous example, we will treat the digitized points around the outlines as landmarks in order to compute centroid size. They are not real landmarks, and will not be homologous from specimen to specimen, but for the purpose of size calculation this is not important. There are other size measures that could be reasonably used in this case - Crônier et al. used area.

After calculating the centroid sizes, we can attempt a univariate cluster analysis in order to see whether the specimens can be reliably assigned to growth stages based on size alone. Using Ward's method, the specimens indeed make reasonably well separated clusters that are congruent with the classification into growth stages.

Bookstein coordinates

When we have digitized landmarks from a number of specimens, we can not assume that all the specimens have been digitized in similar positions and orientations. Also, for shape analysis we are not primarily interested in size (in morphometrics the word 'shape' is even defined such that it excludes size), and therefore we need to remove size as well as position and rotation.

A simple approach to this registration problem is to select two landmarks which will define a *baseline*. Each shape is then translated, rotated and scaled such that these two landmarks are placed at coordinates (0,0) and (1,0), respectively. All the remaining landmarks for each shape make up the so-called *Bookstein coordinates* for the shape. For example, all triangles can be placed in a two-dimensional Bookstein shape space, using the two coordinates of the remaining single landmark after the two baseline landmarks have been standardized to (0,0) and (1,0).

Fitting using a baseline is a simple procedure, but it can be criticized on the grounds that the baseline landmarks are chosen subjectively and will be disproportionately weighted in the fitting process.

Procrustes fitting (superimposition)

Procrustes fitting of a number of shapes involves moving, scaling and rotating all the shapes such that the sum of the distances between corresponding landmarks is as small as possible in least squares (figs 11.2-11.3). The moving and rotating steps will probably give more or less the same result as if we tried to align by eye a number of shapes drawn on transparencies. Procrustes fitting thus treats all landmarks equivalently, avoiding the subjective weighting of two particular landmarks as in baseline fitting.

In two dimensions, Procrustes fitting has a solution given by a formula, and can be computed relatively easily. In three dimensions, such a formula does not exist and the fitting must be optimized by an iterative algorithm.

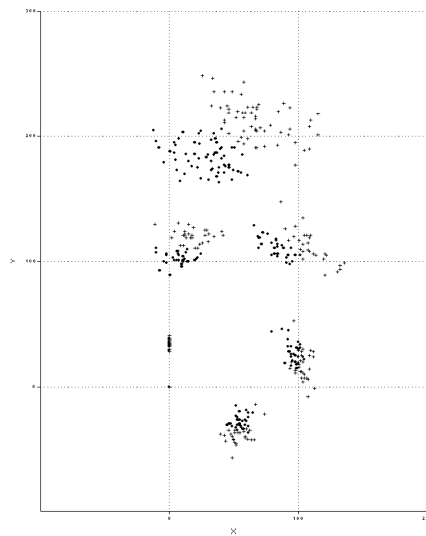


Figure 11.2: Eight landmarks from 51 gorilla skulls (O'Higgins & Dryden 1992).

Procrustes fitting by least squares can sometimes give undue weight to isolated landmarks, such as the tip of a bird's beak. Some attempts have been made to develop robust fitting algorithms which reduce such problems, but they are not in common use.

Analysing landmark-based shapes

A shape defined by a set of landmarks can be viewed as a single point or vector in multidimensional space, where the number of dimensions is two times (for 2D) or three times (for 3D) the number of landmarks.

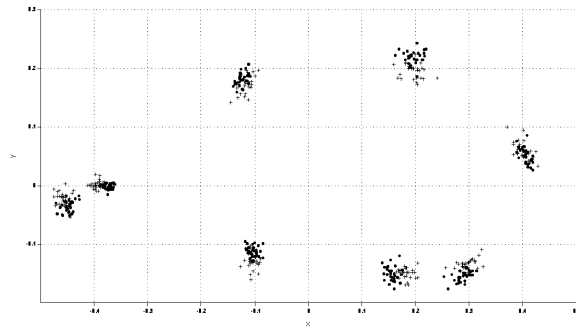


Figure 11.3: The landmarks from fig. 11.2, Procrustes fitted to remove size, rotation and translation.

One issue here, which has produced an enormous amount of literature, is that Procrustes (or Bookstein) fitted shapes are placed in a non-Euclidean shape space. This leads to all kinds of theoretical complications, which we will not discuss further here. In practice, when variation in shape is reasonably small, we can project the shapes into a Euclidean space called the *tangent space*. A quick-and-dirty approximation is to more or less forget the whole issue, and let tangent space coordinates be approximated by the so-called Procrustes residuals, which are simply the Procrustes shapes minus the mean shape.

The Procrustes distance between two shapes is the square root of the sum of squared Euclidean distances between corresponding landmarks in the two shapes.

When given a number of such shapes, we can apply standard multivariate statistical tests on the shape vectors. One such test is Hotelling's T^2 , which is the multivariate analogue of the t test. Given that the shapes are distributed according to multivariate normality, we can use this test directly on the Procrustes residuals to investigate whether two groups of shapes (males/females, two species, two morphs, fossils from two localities) are equal. Similarly, if we have more than two groups, we can use MANOVA to see if they all have the same multivariate mean.

Another option is to perform cluster analysis using the Euclidean distance measure on the set of coordinate vectors. This is equivalent to clustering based on Procrustes distance. Ward's method may be the best choice of clustering algorithm for this application.

PCA of landmark-based shapes

The landmark coordinates of a set of Procrustes-fitted shapes can be directly subjected to PCA. This will produce a set of eigenvectors (principal components), each containing linear displacement vectors for all landmarks. The first principal component will explain more of the shape variation than any other component. Consider an artificial example where corresponding landmarks in all specimens coincide, apart from one landmark which is displaced at different positions along a straight line in the different specimens. There will then be only one principal component (all the others will be zero), and the only non-zero elements in this eigenvector will be the two elements corresponding to the x and y coordinates of the displaced landmark.

With this method, a set of landmark-based shapes can be plotted as points in a low-dimensional space

spanned by the principal components, and trends and groupings within the material be visualized (fig. 11.4).

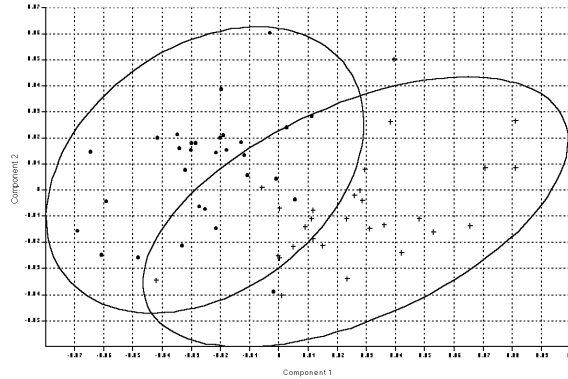


Figure 11.4: Procrustes fitted gorilla landmarks projected onto the two first PCA axes, with 95 percent confidence ellipses. Crosses are males, dots are females.

Also, the landmark displacements corresponding to the principal components can be shown in order to make geometric interpretations. For example, it may turn out that most of the variation in shape is accounted for by the first eigenvector. Investigation of the corresponding landmark displacements may show that this principal component contains variation in a single geometric parameter such as length of the tail or distance between the eyes (fig. 11.5).

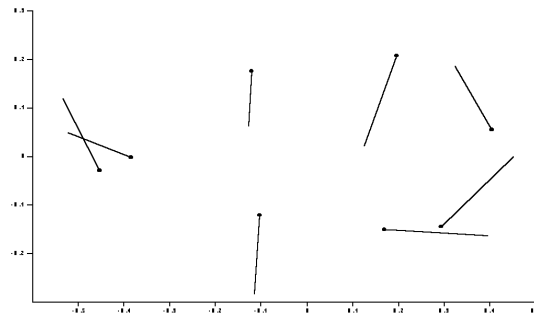


Figure 11.5: Displacement vectors from mean shape (dots), corresponding to the first principal component of the Procrustes fitted gorilla data.

Chapter 12

Thin-plate splines and warps

How can we turn a baboon skull into a chimpanzee skull, or a cod into an eel? D'Arcy Wentworth Thompson famously illustrated such transformations by placing the source shape onto a square grid, and deforming the grid and the shape correspondingly until the target shape was reached. This approach turns out to be very useful, because it gives a clear visualization of shape change and allometries. In addition, multivariate analysis of such grid transformations (also known as warps, or morphings) provides new possibilities for studying variations in shape.

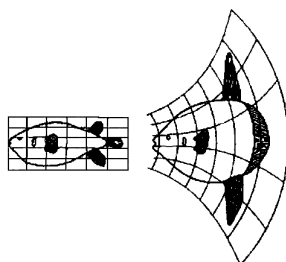


Figure 12.1: One of Wentworth Thompson's original grid transformations (Thompson 1917).

Thompson drew his grid transformations by hand, but we need a more automatic and objective method. We want to find a way of transforming a grid so that it transports the source landmarks onto the destination landmarks, but with a minimum of large-amplitude, small-range twists and compressions.

To understand how this can be done, we will start by looking at a seemingly different, but simpler problem. A number of points are distributed on a thin, flat, horizontal metal plate (fig. 12.2). We want to move the points vertically into a new set of positions. The resulting buckling of the plate can be simulated in a computer using the laws of mechanical physics, and it turns out that the shape of the deformed plate will have a very desirable property: It will be the shape that minimizes total curvature under the constraint of passing through the points. The shape (the thin-plate spline) is a *maximally smooth interpolator*.

Now, we can perform a similar operation independently on landmark displacements in the x and y directions within the paper plane (not vertically as above). This will produce a thin-plate spline that can be used not only to deform the original landmarks from the source to the target, but to move any point in the plane, including the grid nodes. The same principle can be extended to three dimensions.

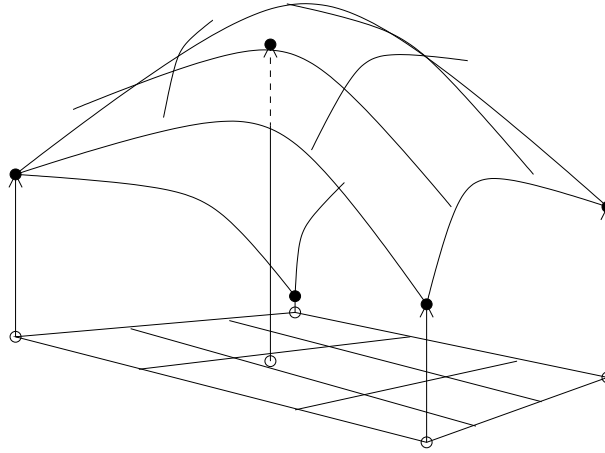


Figure 12.2: Physical deformation of a thin plate, such that five landmarks (open circles) on the original, flat horizontal plate are translated vertically into five new, given positions (filled circles). A grid drawn on the original plate is correspondingly deformed.

Principal warps

As described so far, thin-plate spline theory is comparatively simple from the user's point of view, and it can already be used as a method for visualizing shape change. However, the field has been taken much further by analyzing the deformations in detail and using them as data for multivariate methods. The concepts of *principal*, *partial* and *relative warps* are central here. These powerful extensions have become very popular in morphometrics, but there is little doubt that they can be difficult to understand intuitively, sometimes confusing relatively simple problems that could be studied with more direct methods.

Given a certain configuration of source landmarks, we can construct a 'deformation space', meaning a multidimensional space where all the possible deformations from the one source to any target can be plotted as points. The coordinates in this space can simply be the x and y displacements of each landmark away from the source positions, but it turns out that it is more useful to define an orthogonal set of basis functions for this space based on the 'bending energy' of the transformation. These univariate basis functions are called *principal warps*. I would like to repeat that the principal warps define a space of all possible transformations away from the one given source configuration, and that they are therefore independent of any specific target configuration. In other words, you can look at the landmarks of a single specimen, and compute the corresponding principal warps. A second target specimen is not used for this purpose.

Principal warps are usually only a step in the calculation of partial warps. They are perhaps not very useful in themselves.

Partial warps

The principal warps are only dependent on the source configuration, but the *partial warps* are unique to each transformation from the source to a given target. If we have a collection of specimens, and use the mean shape as the source configuration, each of the specimens will have its own partial warps, dependent

upon the shape change from the mean to that specimen. This means that we can plot all the specimens in a partial warp space.

Any transformation is completely and uniquely defined as a weighted sum of partial warps, in addition to an *affine component* (sometimes called the zeroth-order warp) which is simply the linear component of the transformation. The affine component describes the linear deformations of scaling, stretching and shearing. Hence, the affine component plus the partial warps constitute a complete *decomposition* of the transformation.

The main attraction of the partial warps is that they are ordered according to how local their corresponding deformations are. The first partial warp corresponds to a very large-scale, global deformation, while higher-order warps correspond to more local deformations. The decomposition of a transformation into partial warps can therefore be useful for understanding the biological meaning of the shape change. This ordering according to localization is 'inherited' from the corresponding principal warps, in the sense that the i 'th partial warp can be interpreted as the movement of the landmarks which are most strongly weighted in the i 'th principal warp.

The partial warps are transformations in their own right, and can be visualized with grid deformations (fig. 12.3).

The *partial warp scores* on the other hand, are not transformations but pairs of numbers (one pair for each partial warp) which can be interpreted as the relative importance of the corresponding principal warps in the transformation, in the x and y directions respectively. The partial warp scores for each partial warp can be shown in scatter plots, providing low-dimensional plotting of transformations.

Relative warps

Informally, the relative warps are the principal components of a set of thin-plate spline transformations. The relative warps are themselves transformations, and can be visualized with grid deformations. The original transformations can then be placed as points in a PCA scatter plot (relative warp scores), providing a powerful data reduction technique for a set of transformations. Using the mean shape as the source, the method of relative warps can also be used for the ordination of a collection of shapes.

Growth and allometry versus thin-plate splines

Thin-plate splines are useful for visualizing shape change, but do they have any biological significance? Can, for example, the partial warps be interpreted in terms of growth, allometry and heterochrony? The answer is: only with difficulty.

One major problem is that allometric growth causes non-uniform stretching and rotation which will give different partial warps as ontogeny progresses. Therefore, the partial (or relative) warps can not be compared to see, for example, if two ontogenetic trajectories are similar, or if one species can be viewed as a heterochronic extension of another. Using thin-plate splines for phylogenetic studies is therefore highly problematical.

Even though proper links between thin-plate spline theory and growth mechanisms are yet to be found, there are a couple of simple measurements we can do on the transforming grid which may be of biological significance.

Ideally, we should be able to decompose shape change into an optimal, small set of underlying, orthogonal allometric gradients. As this is as yet impossible, we can at least study the degree of local expansion or contraction of the grid (fig. 12.4). We can either simply measure the area of grid squares before and after the transformation, or we can use a slightly more accurate method based on the partial derivatives of the transforming function (the Jacobian determinant).

This will give us an idea about local growth rates, but will say nothing about directions of growth. For this purpose, a good approach is to draw little circles on the original grid, and study how these are turned into *strain ellipses* by the transformation. The major and minor axes of these ellipses will define the major and minor strains, and will indicate the degree and direction of anisotropic growth. These principal strains can be calculated from the transforming function.

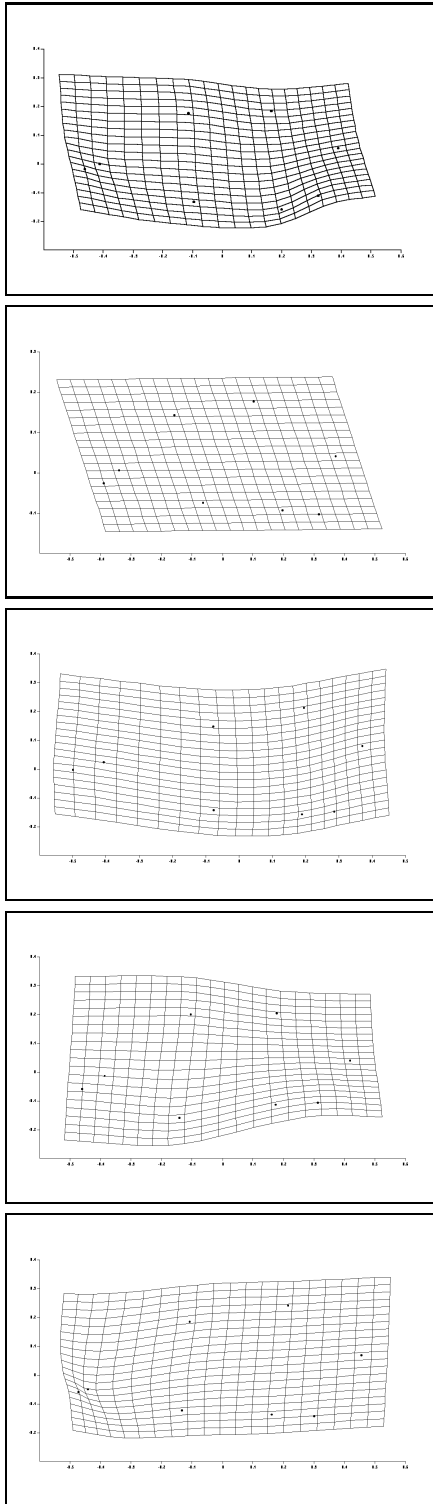


Figure 12.3: Top: Complete warp from female 1 to male 9 in the gorilla data. Below the decomposition into affine component and partial warps 1, 2 and 5 (warps 3 and 4 not shown). Note how partial warp 5 is a purely local deformation, involving only two landmarks.

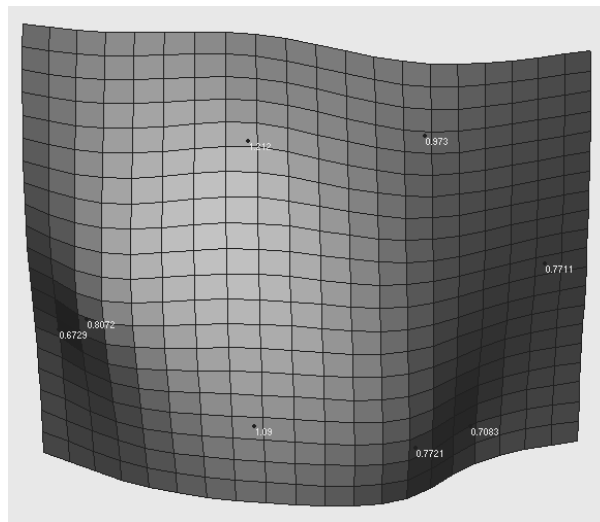


Figure 12.4: A thin-plate spline transformation from a female to a male gorilla skull, with eight 2D landmarks. Expansion or contraction is shown with a gray-scale (output from PAST).

Bibliography

- [1] Bruton, D.L. & A.W. Owen. 1988. The Norwegian Upper Ordovician illaenid trilobites. *Norsk Geologisk Tidsskrift* 68:241-258.
- [2] Cronier, C., S. Renaud, R. Feist & J.-C. Auffray. 1998. Ontogeny of *Trimerocephalus lelievrei* (Trilobita, Phacopida), a representative of the Late Devonian phacopine paedomorphocline: a morphometric approach. *Paleobiology* 24.
- [3] Davis, J.C. 1986. *Statistics and Data Analysis in Geology*. John Wiley & Sons.
- [4] Dryden, I.L. & K.V. Mardia 1998. *Statistical Shape Analysis*. Wiley.
- [5] Ferson, S.F., F.J. Rohlf & R.K. Koehn. 1985. Measuring shape variation of two-dimensional outlines. *Systematic Zoology* 34:59-68.
- [6] Hammer, O. 2000. Spatial organisation of tubercles and terrace lines in *Paradoxides forchhammeri* - evidence of lateral inhibition. *Acta Palaeontologica Polonica* 45:251-270.
- [7] Harper, D.A.T. (ed.). 1999. *Numerical Palaeobiology*. John Wiley & Sons.
- [8] Jolliffe, I.T. 1986. *Principal Component Analysis*. Springer-Verlag.
- [9] Kruskal, J.B. 1964. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika* 29:1-27.
- [10] MacLeod, N. 1999. Generalizing and extending the eigenshape method of shape space visualization and analysis. *Paleobiology* 25:107-138.
- [11] O'Higgins, P. & I.L. Dryden. 1992. Studies of craniofacial development and evolution. *Archaeology and Physical Anthropology in Oceania* 27:105-112.
- [12] Palmer, A.R. 1994. Fluctuating asymmetry analyses: A primer, pp. 335-364. In T. A. Markow (ed.), *Developmental Instability: Its Origins and Evolutionary Implications*. Kluwer, Dordrecht, Netherlands.
- [13] Press, W.H., S.A. Teukolsky, W.T. Vetterling & B.P. Flannery. 1992. *Numerical Recipes in C*. Cambridge University Press.
- [14] Reyment, R. 1991. *Multidimensional Paleobiology*. Pergamon Press.
- [15] Thompson, D'A.W. 1917. On growth and form.