

30 Multivariate analyses of morphometric turtle data — size and shape

Claude, J., Jolliffe, I.T., Zuur, A.F., Ieno, E.N. and Smith, G.M.

30.1 Introduction

Morphometry is the measurement of shape. For most morphometric studies, a large number of measurements is required. For example, in Chapter 14, we used six body measurements collected on 1100 sparrows. One option is to analyse each variable separately using univariate methods. But this is a time-consuming process and, moreover, the multivariate nature of the data is not taken into account. A sensible approach is then to apply multivariate techniques, and obvious candidates are often principal component analysis (PCA) and discriminant analysis (DA). These techniques were discussed in Chapters 12 and 14. DA can be used if there is a grouping in the observations, as we had for the sparrow data (different observers, species or sex). Although PCA was successfully applied on the dolphin data in Chapter 29, there is a problem with PCA if it is applied on morphometric data.

Recall that the first axis in a PCA represents the major source of variation in the data. If all morphometric variables are related to the overall size of an organism, then typically the first axis is determined by size. Typically, all morphometric variables have a similar influence on this axis and for that reason it is often called a size axis. Because further axes are required to be orthogonal to this axis, we end up with axes that mostly describe some aspect of the shape of the subject. For this reason, we talk about size and shape axes. In practice, this often means that the first axis explains most of the variation in the data.

The aim of this chapter is to discuss several methods that can be used to deal with the size and shape problem. We will discuss classic approaches based on pre-standardisation and PCA, but we will also present a more recently developed approach that analyses the relative positions of landmarks that can be digitised on a picture or directly on the biological structure of interest (Bookstein 1991). It is referred to as *geometric morphometrics*, and an increasing amount of effort has been devoted to this approach (Bookstein 1991; Dryden and Mardia 1998; Zelditch et al. 2004). To illustrate the methods, a turtle morphometric dataset is used.

The outline of the chapter is as follows. We will first introduce the data and apply a short data exploration. We will then discuss a series of classic approaches, and we present results of some of these approaches for the turtle data. Finally, an

introduction into geometric morphometrics is given and a short illustration is provided.

Turtles in the dataset are living aquatic and terrestrial species from Europe, Asia, Africa, South and North America. Most of them are housed in the collections of the Chelonian Research Institute (Oviedo, Florida), some are from the National Museum of Natural History of Paris, and some are from Marc Cheylan, Haiyan Tong and Julien Claude. Institutional numbers of specimens are available in Claude et al. (2003, 2004).

30.2 The turtle data

Since their first occurrence in the late Triassic fossil record (225 million years ago), many families of turtles have appeared and have evolved to marine, freshwater and terrestrial environments. Turtles are unique among living amniotes because they possess an anapsid skull. Despite their primitive appearance their skull exhibits a wide array of morphologies; this has been important to morphologists to establish phylogenetic scenarios (e.g., Gaffney 1975).

The skull is an integrated morphology, and any description should take into account relations between the morphological traits that compose its structure. In addition there is a huge size variation between turtles, something that is not necessarily of interest if we want to focus on shape variation.

Few studies have attempted to describe the morphological variation of the skull in quantitative terms (Claude et al. 2004), although these approaches can demonstrate explicitly the role of evolutionary mechanisms such as selection or constraints.

Twenty-four measurements were obtained from inter-landmark distances on the skulls of 123 species of turtles (Figure 30.1). We will denote these variables as D_1 to D_{24} in this chapter. A description of these variables is presented online. The sample covers all the modern families of turtles (14) and the three main ecological groups (freshwater, marine and terrestrial). In addition, the skull of the more primitive and oldest turtle, *Proganochelys quenstedti* (225 million year old), was added to the data. For some species, several individuals were available to correct for intra-species variation. Hence, for some species the observation corresponds to the average of individual measurements. A list of taxa and effective size for each species is available online.

As we will show in Section 30.4, the variation in size among the 123 species is considerable for this dataset. The overall size of a turtle skull is often described by the midline length of the specimen from the premaxilla to the occipital condyle. This is D_2 in Figure 30.1. Its size ranges from 23 mm for the smallest specimen (*Testudo kleinmani*) to more than 220 mm for the biggest species analysed here.

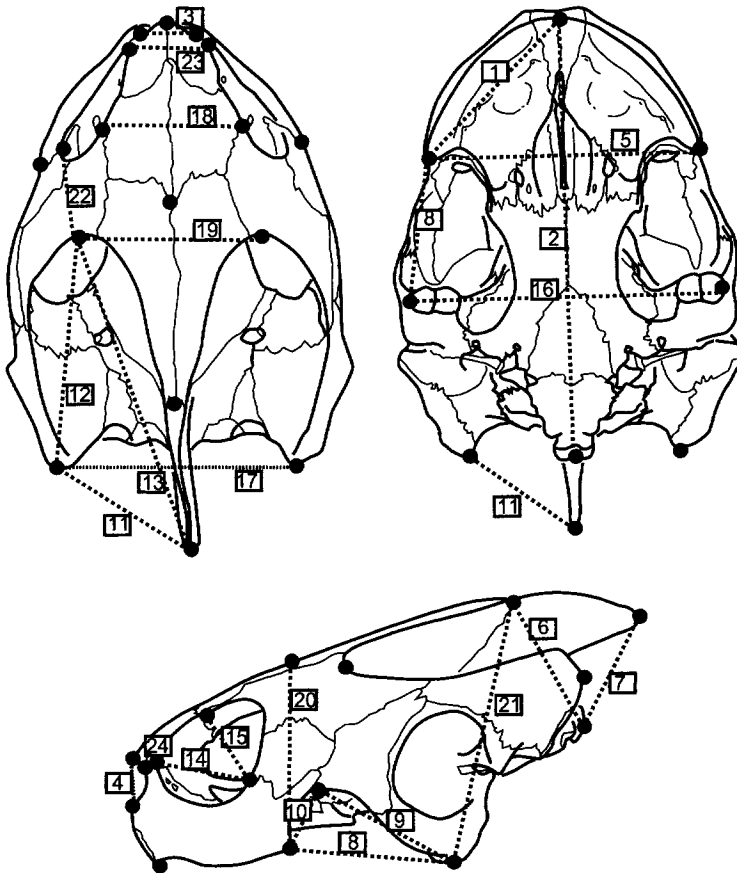


Figure 30.1. Definition of distance measures. The 24 distance variables will be called D_1 to D_{24} in the text. These measures are inter-landmark distance. The 25 landmarks (black dots) will be used in the Section 30.7.

30.3 Data exploration

The first question we address is whether there are any outliers in the data. An outlier is in this case a turtle with an extreme large or small value for one or more distance variables. Graphical tools to identify such observations are the boxplot and Cleveland dotplot; both were discussed in Chapter 4. The boxplot for all distance variables is given in Figure 30.2, and it shows that (i) not all distance variables are within the same range (some vary around 50 mm and others around 10 mm) and (ii) there is considerable variation within each variable. The first point

may sound trivial, especially if you look at how these variables were defined in Figure 30.1. For example, D_3 and D_4 are considerably smaller than D_2 and D_{21} and have therefore smaller spread. However, this immediately raises an important question for the PCA; do we consider all distance variables equally important or do we mainly focus on the variables with larger spread? The answer to this question is based on biology and not on statistics. In this case, it was felt that all distance variables were equally important, and hence, we will use a correlation-based PCA.

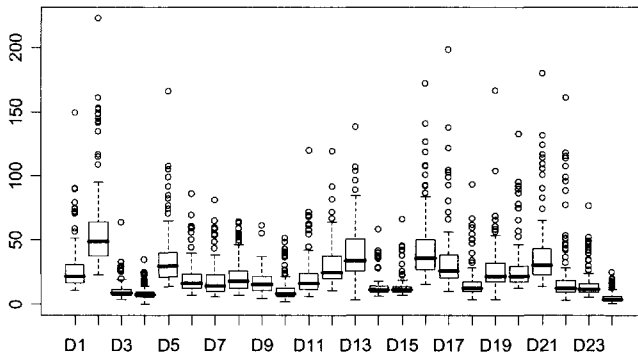


Figure 30.2. Boxplot of the 24 distance variables. Due to limited space along the horizontal axis, not all labels are printed.

To answer the question, whether there are outliers, Cleveland dotplots (not presented here) confirmed the impression already given by the boxplots, namely that some variables would benefit from a transformation. The reason for this is that for some variables there were individual observations, and also groups of observations, that were relatively far away from the centre of the data. Before applying any data transformation, we constructed a pairplot (Figure 30.3), which shows that these observations are not real outliers as they comply with the linear relationship between the variables (see also Section 4.3). For this reason we did not apply a transformation on the distance variables. Note that most of the correlation coefficients (Figure 30.3) between the distance variables are relatively large. This is typical for morphometric data; a turtle with a large skull tends to have large values for all distance variables. Another possible reason for the high correlation coefficients can be a grouping structure in the observations. This is often the case when biological groups segregate the observations. In our example, some turtle species share a similar way of life that may generate resemblance because of convergent evolution. In addition, species belong to different groups (genera, families, superfamilies) that have distinct evolutionary history, and this will directly result in segregation among the observed morphologies. Imagine, for example, that we have

two morphologically distinct families in our data. This will produce a correlation structure in the data, even if within a family, variables are unrelated. To illustrate such a grouping structure (and associated problems), suppose we have *artificial* data with values:

Y_1 : 1 2 1 3 4 1 2 100 99 80
 Y_2 : 2 3 4 1 6 3 1 96 90 75
 Y_3 : 3 2 6 6 3 4 9 99 60 84

Biologically speaking the higher values for the three last observations for the three variables may have evolutionary grounds. As an example, imagine that the three last observations belong to a different taxonomic group (family A versus family B) or to a distinct ecological group (aquatic versus terrestrial forms, for example). For the turtle data we do not have Y_1 to Y_3 with 10 observations each, but D_1 to D_{24} with 123 observations each. However, the following discussion is relevant for the turtle data. The correlation coefficient among Y_1 , Y_2 and Y_3 will be close to 1 only because a group of observations has much higher values. We have similar problems in the turtle data; turtles from particular environments have all higher distance values (this can be illustrated with, for example, a Cleveland dot-plot).

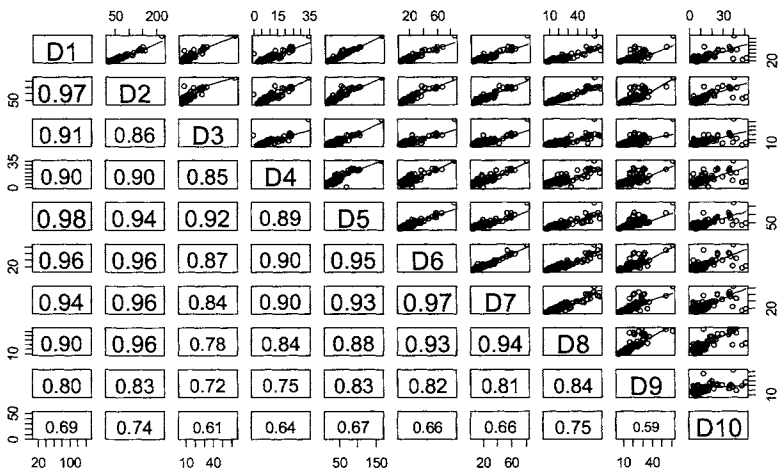


Figure 30.3. Pairplot for selected distance variables (D_1 to D_{10}). The lower panel contains correlation coefficients between distance variables. The font size is proportional to the correlation.

Dependence among observation is more difficult if we want to take phylogeny into account. In this case we have to take into account the shared evolutionary history of pairs of species that will be expressed as the amount of independent morphological divergence. As species sharing a more recent common ancestor are

more likely to look similar, the dependence structure among observations can be directly related to the phylogenetic tree structure. Generalised estimating equations (GEE, an extension of GLM) have been used in this context (Paradis and Claude 2002). In addition to GLM, GEE introduces a correlation structure among observation. Paradis and Claude (2002) have explained how a phylogenetic tree can be transformed into this correlation matrix. However, this issue is beyond the scope of this chapter. Other methods have been attempted to correct the phylogenetic effect (Felsenstein 1985; Cheverud et al. 1985; Grafen 1989; Martins and Hansen 1997), following diverse approaches (auto-regression, “contrast” analysis, generalised least squares).

30.4 Overview of classic approaches related to PCA

There is a large literature on the study of size and shape using morphometric data, and various approaches have been suggested for quantifying size and shape. Several of them use PCA-related techniques and we concentrate on these in this section, which is based on Section 13.2 of Jolliffe (2002). By ‘classic’ we mean that inter-landmark distances are analysed, rather than the co-ordinates of the landmarks themselves. Analyses based on the latter are discussed in Section 30.7.

The idea of using the first principal component (PC) as a measure of size, with subsequent PCs defining various aspects of shape, dates back at least to Jolicoeur (1963). It is fairly conventional, for reasons explained in Jolicoeur (1963), to take logarithms of the data, with PCA then conducted on the covariance matrix of the log-transformed data.

Whether PCA is done on untransformed or log-transformed data, the coefficients or loadings for all variables in the first PC will often be similar, but they will not be identical. Some authors argue that all variables should have *equal* importance in measuring size, so that the first PC should be replaced by an equally weighted combination of all variables. This is known as *isometric size*. Somers (1989), for example, argues that the first PC contains a mixture of size and shape information, and that in order to examine ‘shape’, an isometric component rather than the first PC should be removed. Several ways of removing isometric size and then quantifying different aspects of shape have been suggested, some of which are discussed below. The problem is that by replacing the first PC by isometric size, at least one desirable property of PCA is lost.

It does not help that none of the terms size, shape, isometric or allometry is uniquely defined, which leaves plenty of scope for vigorous debate on the merits or otherwise of various procedures; see for example Bookstein (1989) or Jungers et al. (1995).

PCA has the two properties that vectors of loadings defining the PCs are orthogonal and that the PCs themselves are uncorrelated. Both of these properties are desirable as they mean that, in two senses, different PCs are measuring clearly different things. Most alternatives that attempt to use the isometric size instead of

the PCs lose at least one of these two properties. The covariance matrix S of a dataset consisting of measurements on p variables x , can be written as

$$S = l_1 a_1 a_1' + l_2 a_2 a_2' + \dots + l_p a_p a_p'$$

where $l_1 \geq l_2 \geq \dots \geq l_p$ are the eigenvalues of S , and a_1, a_2, \dots, a_p are the corresponding eigenvectors. If the first term on the right-hand-side of this equation is removed and a PCA is done on the resulting matrix, the first PC is the second PC of S , the second PC of the new matrix is the third PC of S , and so on. Somers (1986) suggested removing $l_0 a_0 a_0'$ from S and doing a 'PCA' on the resulting matrix to get 'shape' components, where l_0 is the variance of $a_0'x$ and

$$a_0 = \frac{1}{\sqrt{p}}(1, 1, \dots, 1)$$

This procedure has several drawbacks (Somers 1989; Sundberg 1989). In particular, the shape components lose the desirable properties mentioned above.

One alternative suggested by Somers (1989) is to find shape components by doing a PCA on a double-centred version of the log-transformed data. Double-centring subtracts the mean of each row in a dataset as well the mean of each column, whereas ordinary PCA only centres the columns (variables). It can be considered to remove size because the isometric vector is one of the eigenvectors of its covariance matrix, but with zero eigenvalue. Hence the vectors of coefficients of the shape components are orthogonal to the isometric vector, but the shape components themselves are correlated with the isometric component. Cadima and Jolliffe (1996) quote an example in which these correlations are as large as 0.92. Double-centring is illustrated for the turtle data in Section 30.6.

By using projections, Ranatunga (1989) formulated a method for which the shape components are uncorrelated with the isometric component, but her technique sacrifices orthogonality of the vectors of coefficients.

Once we move away from eigenvectors of the original covariance or correlation matrix, preserving both properties becomes difficult. Cadima and Jolliffe (1996) succeeded in deriving a procedure, combining aspects of double-centring and Ranatunga's approach, that gives shape components that are both uncorrelated with the isometric component and have vectors of coefficients orthogonal to a_0 . However, as pointed out by Mardia et al. (1996), another fundamental property of shape components is lost in this method. If $x_h = cx_i$, where x_h, x_i are two observations of the p variables and c is a constant, then most definitions would say that x_h, x_i have the same shape. However, the scores of the two observations are generally different on the 'shape components' found by Cadima and Jolliffe's (1996) method.

Another approach to the analysis of size and shape is to define a scalar measure of size, and then calculate a shape vector as the original vector x of p measurements divided by the size. This is intuitively reasonable but needs a definition of size. Darroch and Mosimann (1985) list several possibilities but hone in on:

$$g_a(x) = \prod_{k=1}^p x_k^{a_k}, \quad \text{where } a = (a_1, a_2, \dots, a_p) \quad \text{and} \quad \sum_{k=1}^p a_k = 1$$

So size is thus a generalization of the geometric mean. We now have a shape vector $x/g_a(x)$ but may want to decompose it into scalar shape components. This brings us back to PCA. Darroch and Mosimann (1985) suggest using PCA on the log-transformed shape vector, leading to shape components. They show that the PCs are invariant with respect to the choice of a , provided that its elements sum to 1. The covariance matrix of the log shape data has the isometric vector as an eigenvector, with zero eigenvalue, so that all shape components are contrasts between log-transformed variables.

30.5 Applying PCA to the original turtle data

To illustrate what happens if PCA is applied on morphometric data, we apply it to the turtle data. As discussed in Chapter 12, two main questions have to be addressed before applying PCA, namely (i) should all the distance variables have the same influence or should the variables with more variation dominate the analysis, and (ii) are we interested in relationships between observations (turtles) or between distance variables? The first question is directly related to the problem of whether we should apply the PCA on the correlation or covariance matrix, and the second question dictates whether we should use the distance biplot or (confusingly) the correlation biplot. The correlation biplot is also called the species conditional biplot, but that would cause even more confusion, as the turtle species are actually the observations.

As noted in Section 30.3, we decided to use the correlation matrix and correlation biplot. As a result we obtain a biplot in which angles between lines (the distance variables) represent correlations between the distance variables, but distances between observations (turtles) cannot be compared directly with each other; they are approximate Mahalanobis distances that can sometimes be a bit more difficult to interpret. This means that if we obtain the biplot, we should not enthusiastically start to compare different groups of observations (e.g., freshwater and marine turtles) as if distances were Euclidean, but merely concentrate on the distance variables.

The biplot is given in Figure 30.4. Some software packages scale the eigenvalues so that the sum of all eigenvalues is equal to 1, and we use this convention here. The first two eigenvalues are 0.84 and 0.07, respectively, which means that the first two axes explain 91% of the variation. All variables are strongly related to the first axis. This is typical for morphometric data, and the first axis can be thought of as a size axis (Jolliffe 2002).

For most datasets we would be very content with 91% explained variation, but not in this case. Recall from Chapter 12 that this 91% means that the biplot explains the correlation matrix rather well, but the correlation matrix itself is mainly determined by length of the skulls. So the first axis represents a rather trivial outcome, namely that most distance variables are highly correlated.

In the next section we look at several alternatives to straightforward PCA, using the inter-landmark data.

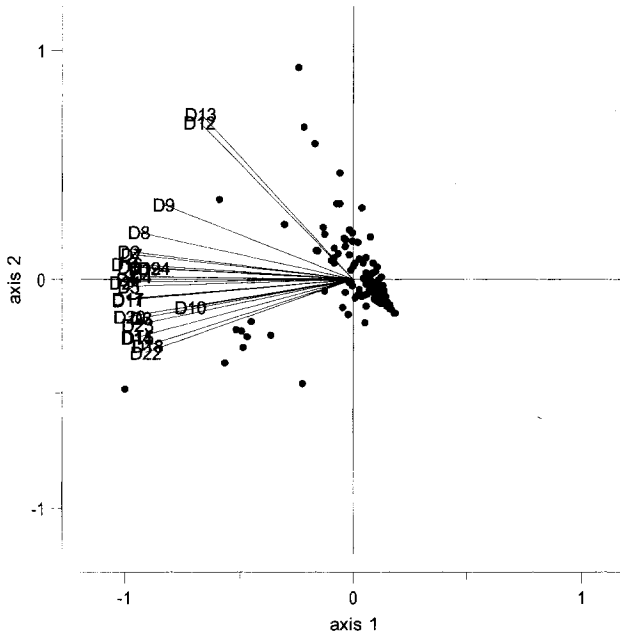


Figure 30.4. PCA biplot; axis 1 versus axis 2. The correlation biplot (also called: species conditional) biplot was made based on the correlation matrix. The first three eigenvalues are 0.84, 0.07 and 0.03, respectively. Eigenvalues were scaled to have a sum of one, which means that the first two axes explain 91% of the variation in the data.

30.6 Classic morphometric data analysis approaches

As noted earlier, by classic we mean analyses based on inter-landmark distances, as opposed to those based on the co-ordinates of the landmarks.

Option 1: Using variable D_2 as a scalar measure of size

The first option we discuss is filtering out the size effect by using a scalar measure of size. For some morphometric datasets, an overall measure of length is available. For the turtle data, the variable D_2 may be considered as the overall measure of size, but this is an arbitrary choice specific to this dataset only. To filter out the overall-size effect, it may be an option to apply a linear regression on each distance variable (except for D_2) and use D_2 as the explanatory variable. The

residuals of these regression models are uncorrelated with the general size variable D_2 . This process was applied on the turtle data. We used variable D_2 as a proxy for overall length of the skull, although this is a choice that can be argued about. The PCA correlation biplot (based on the correlation matrix of residuals from the procedure) is given in Figure 30.5. Note that we no longer have the situation in which all variables point in the same direction.

The first axis represents 46% of the variation and shows distinctions between robust skulls with shallow temporal emargination and skulls with deep emargination (see the opposition between D_{12} , D_{13} and most other distances on the first shape axis). The second axis accounts for 15% of the variation and is also influenced by the degree of temporal emargination (D_{12} , D_{13} , D_{22}), the development of the length of the posterior part of the skull (D_8 , D_9), the location of the jugal emargination (D_{10} , D_9), and the relative development of the orbit (D_{14} , D_{15}). However most observations are grouped close to the origin and this first attempt to remove size from shape seems to yield rather little biological information in terms of the structure of the shape space. Allometry (the relationship between size and shape) may drive such a pattern, at least on PC1, because a lot of variables are contributing in the same way to this axis. As for the observations (the turtles), the axes seem to depend greatly on peramorphic (hyper-adult) morphologies of giant species that represent often extreme plots (the narrow-headed giant softshell — maximal score on axis 1— and the giant leatherback turtle —minimum score on axis 1—, the alligator turtle —minimum score on axis 2—). As these species are among the biggest in our sample, this may reflect different allometric relationships. Only groups like marine turtles and trionychid turtles were distinct from each other on these two axes. However, for a statement like this, it is better to use a distance biplot and not a species conditional biplot.

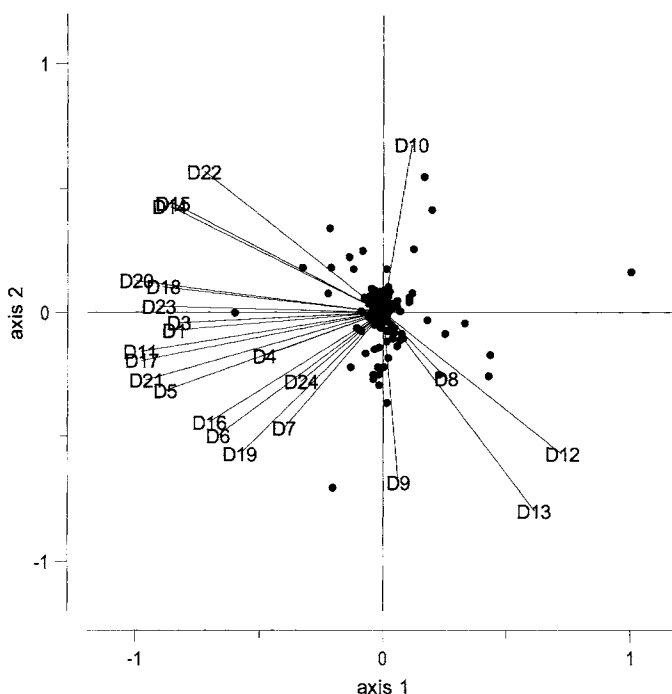


Figure 30.5. Correlation biplot for the data in which D_2 was partialled out. The first three eigenvalues are 0.46, 0.15 and 0.09.

Option 2: Removing isometric size with a pre-standardisation

The next option we discuss is a pre-standardisation of each observation with a size estimator. We choose to use the total squared distance of each observation from the origin as a general measure of size for each observation. This means that for each species, we calculate

$$D_{tot} = \sqrt{\sum_{j=1}^{24} D_j^2} \quad (30.1)$$

The variable D_j is the original distance variable (no transformations were applied on it). Then each of the 24 observations for this species is divided by D_{tot} . This process is repeated for all 123 species. After this pre-standardisation, each distance variable is normalised so that all variables have the same weight in the PCA. In this case, we use the distance biplot as we are interested in how the turtles are projected on the two-dimensional space. Compared with the previous analysis the first two axes contain less variation (39%), and the eigenvalues decrease less rapidly; the first three eigenvalues are 0.24, 0.15 and 0.13. The biplot is given in Figure 30.6-A and Figure 30.6-B. We used two panels, one for the variables and

one for the scores as we did not want to clutter the graphs too much. So, one has to imagine these two panels on top of each other. As this is the distance biplot, we cannot compare the distance variables by their angles. All we can say is that variables pointing in the same direction have similar values for certain groups of turtles. Panel A shows that we can indeed see a certain grouping in the variables, namely (i) D_8, D_9 , (ii) D_{12}, D_{13} , (iii) $D_3, D_4, D_{14}, D_{15}, D_{18}, D_{20}$ and D_{23} , and (iv) D_{11}, D_{17} and D_{22} . Some of the variables were poorly represented on the first two axes (e.g., $D_2, D_{16}, D_{19}, D_{24}$). In panels B and C we used super family and environment information as labels. Groups of observations corresponding to either taxonomic families or ecologies appear clearly. The first axis seems to represent a distinction between marine and terrestrial turtles versus all the other freshwater turtles. This axis represents, among other things, a change in the relative middle height of the skull (D_{20}) and relative size of the posterior part of the skull (D_8, D_9), which varies from flat and triangular with more developed posterior part versus higher, more rectangular and shorter posterior part in lateral views along the axis. The second axis opposes marine turtles (bottom) to terrestrial ones (top). Species close to the top of the graph get a shortest relative distance between the occipital condyle and the supraoccipital crest (D_6, D_7), longer distance between supraoccipital crest and ends of squamosals (D_{11}), deeper temporal emargination (D_{12}, D_{13}), longer basioccipital to maxillary length (D_2), and wider orbits and nares (D_3, D_4, D_{14}, D_{15}). Clades were grouped rather well on the first axes; for example, trionychoids (labelled 3) were gathered at the top-left part of the plot, representing flat and deeply emarginated skulls.

The pre-standardisation carried out above was also discussed in Chapter 12. Distances between observations (turtles) in the PCA biplot (applied on the transformed data) are now two-dimensional approximations of the Chord distance (Chapter 10) between two turtles (Legendre and Gallagher 2001).

Option 3: Removing isometric size with double centring

There is another pre-standardisation option, namely double centring, which was described in Section 30.4. If we imagine the data as being from 10 turtles (T_1 by T_{10}), we have:

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Y_1 :	1	2	1	3	4	1	2	100	99	80
Y_2 :	2	3	4	1	6	3	1	96	90	75
Y_3 :	3	2	6	6	3	4	9	99	60	84

Turtles T_8, T_9 and T_{10} are just large species. Because of these three, the correlation coefficients among Y_1, Y_2 and Y_3 are very high. To down-weight their effect on the correlation coefficients (and therefore in the PCA), we could calculate the mean value for each turtle over the three observations (the mean for T_1 is 2, etc.), and subtract this mean value from the corresponding column. For the artificial data, this gives

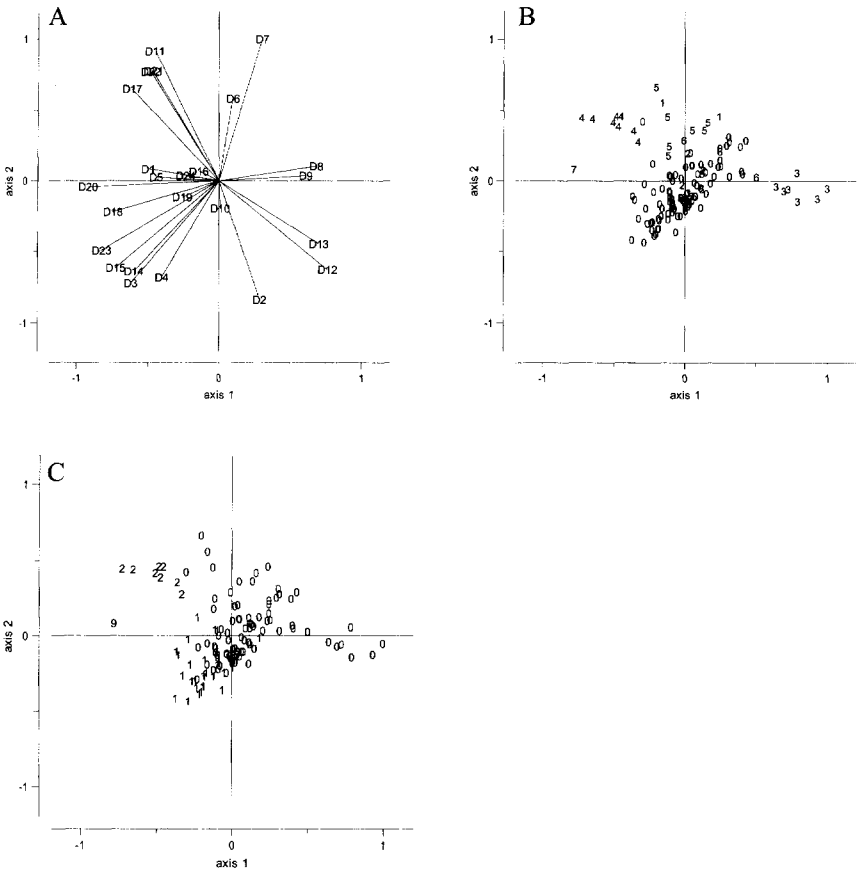


Figure 30.6. A: Axes 1 and 2 for the PCA distance biplot. Only the variables are shown. B: Scores of the PCA distance biplot in panel A. The numbers refer to super-families, and the ‘7’ is the fossil turtle. C: Scores of the PCA distance biplot in panel A. The numbers refer to the environment (0 = freshwater, 1 = terrestrial, 2 = marine and 9 is the fossil turtle).

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
Y_1 :	-1.00	-0.33	-2.67	-0.33	-0.33	-1.67	-2.00	1.67	16.00	0.33
Y_2 :	0.00	0.67	0.33	-2.33	1.67	0.33	-3.00	-2.33	7.00	-4.67
Y_3 :	1.00	-0.33	2.33	2.67	-1.33	1.33	5.00	0.67	-23.00	4.33

The data have been centred by columns. Calculating correlations or covariances, which is necessary for PCA, for these column-centred data will automatically centre by rows — hence the name double-centring. We are slightly inconsistent here as double centring in Section 30.4 was explained for data in which the

columns represent the variables, whereas in the artificial example above, the variables are in rows (to save space).

Results of the double centring are given in Figure 30.7. Observations on each turtle were first centred, and then a PCA using the covariance matrix was applied. The turtles' scores are plotted as labels using the environment in which they lived (0 = freshwater, 1 = terrestrial, 2 = marine and 9 is the fossil turtle). Note that the marine species are grouped together and that these have higher values for D_2 , D_{17} and D_{22} and lower for D_{12} and D_{13} (implying a shallow to absent temporal emargination).

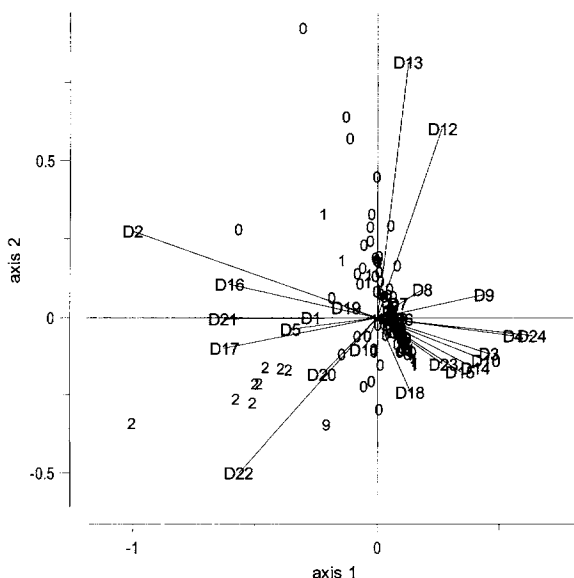


Figure 30.7. PCA biplot based on the covariance matrix after double centring the data of each turtle. The labels for the turtles refer their environment (0 = freshwater, 1 = terrestrial, 2 = marine and 9 is the fossil turtle). The first two eigenvalues are 0.62 and 0.22 corresponding to 62% and 22% of the total variation, respectively.

30.7 A geometric morphometric approach

In the previous analyses, shape components were extracted directly from measurements. These measurements were defined as inter-landmark distances. We have seen that once the size effect is removed, the relationships between these corrected measures and the position of individuals on principal component biplots can be ideally interpreted as shape differences. It is legitimate to think that these shape differences represent, more or less, the relative position of landmarks in

each configuration (a configuration here is defined as a set of landmarks). Imagine that rather than taking many measurements, we had scaled, rotated and translated all the configurations on their centroid (the position of the averaged coordinates of a configuration). Differences between configurations can be quantified as differences of position between corresponding landmarks on an optimally superimposed configuration. Indeed shape information is what we have to retain when translation, rotation and scale effect are removed from the original data.

Geometric morphometrics (introduced in the 1970s) are a set of statistical tools devoted to the study of shape variation. They are particularly relevant for isolating the size and shape components from a configuration of landmarks. The more common methods use superimposition procedures to fit the set of configurations, followed by statistical methods applied to the superimposed coordinates of the landmarks of the different configurations. The differences in landmark positions among superimposed configurations reflect shape differences.

The coordinates of 25 landmarks (Figure 30.1) were digitised in three dimensions on the same sample. Each species corresponds to a configuration of 25×3 coordinates. To obtain a shape space, the information about position, size and orientation have to be removed from the original data. For this aim all configurations are superimposed following three steps: The first scales all configurations to the same centroid size (the centroid size is the square root of the sum of squares from the centroid to each landmark of a configuration). The second step translates all configurations to have a common centroid, and the last step rotates all configurations such that distances between corresponding landmarks are minimised (see Figure 30.8). These distances between corresponding landmarks are minimised during the superimposition.

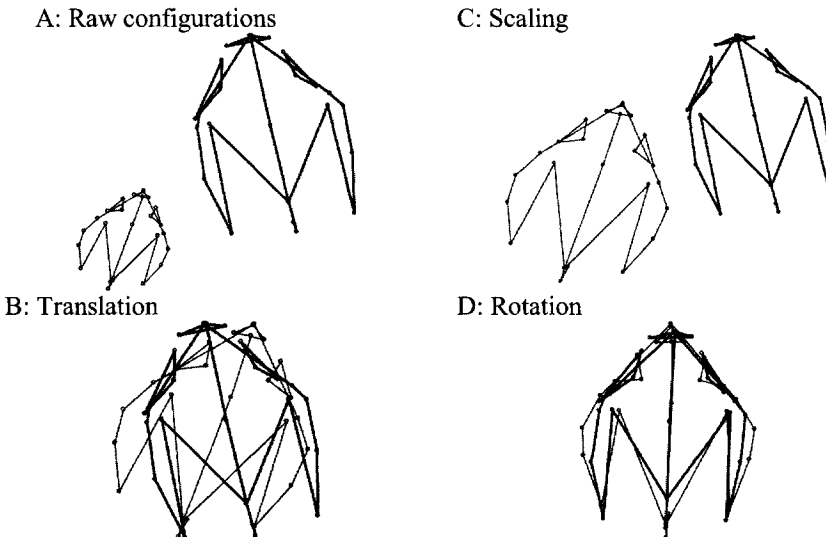


Figure 30.8. The three steps of the superimposition procedure.

For two configurations the procedure is simple. However when there are more than two configurations, finding the best fit uses an iterative procedure: All scaled and centred configurations are first rotated on a reference configuration (for example the first), an average configuration is computed and the rotations are recomputed iteratively until the reference (the average) is stabilized. The details of the transformation are fairly complicated and involve a projection of the shape space (which is non-Euclidian) to a tangent Euclidian one. The interested reader should consult Dryden and Mardia (1998).

Once the superimposition and projection are computed, the resulting transformed coordinates can be used as variables in statistical analysis. As the shape space is multidimensional, principal component analysis on the variance–covariance of coordinates is a way visualizing the position of each shape in this space. Readers interested in other aspects of superimposition methods and shape statistics should again consult Dryden and Mardia (1998).

A shape change along each principal component can be visualised in using the original contribution of each coordinates variables that can be added or subtracted to the average configuration (the origin of the PCA). The score of individuals (here species) can be used for scaling the contribution in change; it is even possible to amplify the eigenvector contribution if shape differences along the axis are not obvious. By adding a few graphical links between landmarks of these extreme shapes, the variation in shape accompanying each axis can be easily described biologically. An illustration is given in Figure 30.9 and Figure 30.10; one illustrates the relation between environments and shape variation and the other the relation between shape variation and super-families.

The first two components accounted for 32.9% and 19.2% of the total shape variance. The different super-families and ecologies are well grouped on the two first PCs, indicating that the geometric shape space directly reflects biological differences. The first component opposes skulls with a deep temporal emargination to skulls with a shallow one. Marine turtles and terrestrial turtles occupy rather opposite locations on this component. Freshwater turtles are more variable, but most of them are close to the origin of the plot. The second principal component opposes high skulls to flatter ones. The flatter skulls have orbits that with a long posterior part of the skull occupy the bottom of the plot, whereas the maxillary part of the skull is more developed and orbits are positioned more laterally for individuals occupying the top of the plot. This PC opposes snapping turtles (chelids, trionychids) to terrestrial ones, with other aquatic and marine ones having an intermediate position. Based on the pattern of variation along two axes, a functional interpretation can be given: The pattern observed on the first axis can be related to skull retraction under the carapace, which is likely related to patterns of temporal emargination — turtles located in the right part of the plot (marine turtles and chelids) have, respectively, no skull retraction, or horizontal skull retraction, whereas those in the left part have a vertical skull retraction. The pattern observed on the second axis can be related to feeding mode — a more elongated anterior part, and shortest posterior part of the skull, is observed in terrestrial turtles that do not need to swallow food as aquatic ones do. See also Claude et al. (2004) for a more detailed analysis. Amazingly, the fossil *Proganochelys* was rather close to

the top of the plot (close to terrestrial and marine turtles), which may be interpreted as resulting from a terrestrial feeding mode, that is, a non-retraction possibility of the neck. We may reach similar interpretations based on the approach used in Section 30.4, option 2. But patterns of skull variation are more easy to interpret with the geometric approach.

Geometric morphometrics seemed to be at least as efficient as the approaches to identify patterns of biological variation described in Section 30.4. In using these two methodologies, outliers influenced PCs less than for the previous options. For geometric morphometrics, this is probably due to the superimposition procedure that spreads variation among the different landmarks of the configurations.

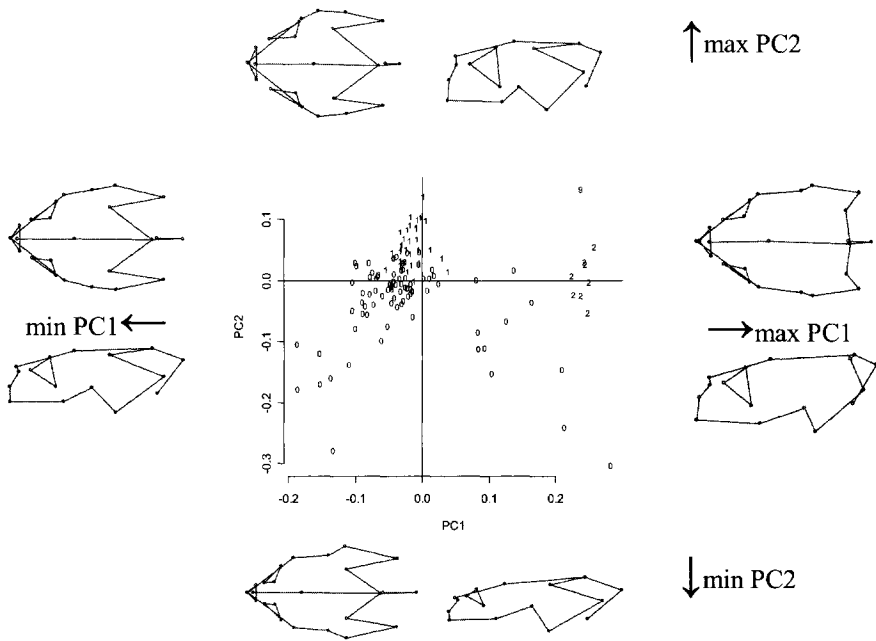


Figure 30.9. Axes 1 and 2 for the PCA of superimposed coordinates. Drawings around the plot represent extreme shapes on each PC. (0 = freshwater, 1 = terrestrial, 2 = marine and 9 is the fossil turtle).

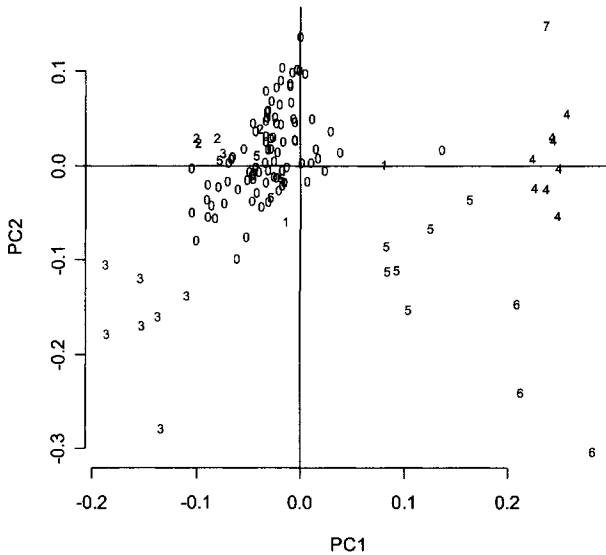


Figure 30.10. Axes 1 and 2 for the PCA of superimposed coordinates. The numbers refer to super-families and the '7' is the fossil turtle.

Many other geometric approaches of shape variation have been developed during the last 30 years: Some methods rely on thin plate spline and deformation grids, others on Fourier analysis, other on angles in configuration, etc. The scope of this chapter is just an introduction; therefore, the reader is invited to read the literature (indicated in the text) if (s)he wants to set a protocol of measurement for analysing shape variation. Geometric morphometrics are now an important alternative approach to a statistical analysis of *ad hoc* inter-landmark distances, and they deal with the size effect efficiently. However, as for any other methods, they can introduce biases and additional problems, the more important one being probably the size estimation that is still subjective.

Acknowledgements

The authors are grateful to P.C.H. Pritchard, H. Tong, M. Cheylan, and F. Renous for providing the material used in this study. This is publication item 2006-046 for J. Claude.