Integrative Biology 200B
"Ecology and Evolution"

University of California, Berkeley
Spring 2011
Updated by Nick Matzke

Lab 6: Intro to Geometric Morphometrics

Introduction

In this lab, we hope to show you how to derive useful variables from geometric morphometrics, by giving you a surface understanding of the theory behind it and experience working with morphometrics software. If you are interested, a valuable morphometrics resource is Jim Rohlf's website at Stony Brook University (http://life.bio.sunysb.edu/morph). The page includes an archive of software (including one program we'll use today), as well as lists of upcoming meetings, a bibliography of morphometrics resources, a glossary of terms, and contact information for people working on morphometrics.

Morphometrics is the mathematical study of shape. There are many times that we are interested in shape for biological studies. For example, when studying allometry or heterochrony, it is often the occurrence of specific shapes relative to other factors such as size or age that is at question. Many of you cold include an analysis of shape in your project. It would be very nice if we could put numbers on shapes and compare those numbers to each other for the statistical analysis of hypotheses about changes in shape.

To study shape first we must define it. We all have an intuitive sense of what the word means, but mathematics requires a more precise definition. According to Kendall (1977) shape is "all the geometric information that remains when **location**, **scale** and **rotational effects** are filtered from an object." As a practical matter we can only study shape by comparing one object to another. That means we have to take a bunch of information about the physical distribution of the different parts of multiple organisms, and then move them into the same place, make them the same size and rotate them, so that they all line up without changing the relative positions of the parts of the organism.

Shape space is a critical concept for understanding how this can be accomplished. The idea is that you take an object and identify the coordinates of a number of homologous points on that object called **landmarks**. That entire object can now be mapped as a single point into a space with as many dimensions as the number of landmarks multiplied by the number of coordinates describing each landmark. For example imagine a line between two points on a two dimensional graph. You could describe that line by assigning an x and a y coordinate to each of those points, giving you four total coordinates. Those four coordinates could now be graphed as one point in a 4-dimensional space. For a triangle you would add one more point and thus you would need a 6-dimensional space to graph it as a single point. If you recorded your original coordinates in three dimensions (x, y and z) you would require a 6D space for a line and a 9D space for a triangle. After mapping this object in our n-dimensional space we transform that space removing position, size and rotation and in the process eliminating several dimensions. What you are left with is shape space.

TPSDIG

The first program we will use today is Jim Rohlf's *TPSDIG*. This is only one of several programs Rohlf has developed, but it is critical for this lab, because it allows us to enter the data we will be analyzing. We use this program to identify the landmarks on our specimens. Landmarks should be homologous points on our organisms that are consistent, repeatable, and coplanar (i.e., they are all visible in a photograph of a specimen). There is no right number of landmarks, but they should be numerous enough to adequately cover and describe a specimen's shape. The details of what a landmark is and how to choose them is a contentious issue that we will avoid for now.

TPSDIG allows us to define landmarks and save them in a format that the other programs can recognize. To use *TPSDIG*, we need to have photos of all of our specimens in the same view, and it is best if all of the photos have a scale. To save time today, we'll just do two specimens.

First let's take a look at the files that this program uses. Open the folder c:/Documents and Settings>All Users>Start Menu>Programs>IB200A> TPS>Examples. This folder contains three files. Two of them are TIF files with the images of fly wings that we will be analyzing. The third is a TPS file called **test**. Open **test** is a word processor. This file points to the two images, which we will use in our analysis, so that they can be analyzed at the same time. As you can see, both of the pictures in this folder are listed as images.

Open **TPSDIG** (Start – Programs) and go to the file menu. Go to **input source** and choose tps file 'test' in c:/Documents and Settings>All Users>Start Menu>Programs>IB200A> TPS>Examples. Both images are from the left wing of female *Aedes canadenses*.

The image file should appear on your screen. Go to the tool bar and select the **bulls'-eye tool**. You can zoom in on the images using the + **button**. Each time you click with the **bull's-eye** a red dot appears representing a landmark. Thus, click on places you want to include as landmarks. Remember that landmarks need to be homologous, so pick points that you can clearly identify in both images. In an ideal situation we would have a scale bar in the image, which you could use to define a standard distance by switching to set scale mode, clicking on the endpoints and entering the length, but we don't. If you make a mistake, click on the pointer to switch to edit mode and move your landmarks around. Select '**Label landmarks**' from the options menu to see in what order you made your marks.

For the next image click the **big red arrow**. Now click on the homologous landmarks in this image in the same order. If you forget where you put a landmark you can use the big red arrows to toggle back and forth between the two images.

When you are done, go to the file menu and select save data as. Save your document to the desktop. Type '.tps' at the end of the file name (For some reason the program can't handle this itself). You can open your file in notepad. If you do, you will see it contains the landmark coordinates for your specimens in raw form.

The rest of the programs we'll be using today were developed by David Sheets at Canisius College. They are already on our computers, but they can downloaded at: www2.canisius.edu/~sheets/morphsoft.html. These are probably the most straightforward and easy to use programs available for morphometric analyses.

CoordGen

The first thing we'll need to do before analyzing our data is match all of our specimens up in shape space. In other words, we need to remove differences in location size and orientation from our data. The methods that allow us to do this are called superimposition methods because we are essentially superimposing all of our data on top of each other. CoordGen can use several superimposition methods, although we will focus on two, two-point registration ('Bookstein coordinates) and procrustes.

Two-point registration is the easier of the two methods to understand. Two points are chosen and a line is drawn between them, making a baseline. All the images are then moved, rotated and changed in size, so that their baselines line up exactly. As a consequence you remove 4 dimensions from our original sets of coordinates, two for each of those points.

Open *CoordGen6d* (the application file!). The path is C:\imp\bin\win32. Keep this window open because we'll return to it for other programs and files. When open CoordGen6f, two windows will appear. The one you want to interact with should be obvious.

In the **tangerine** colored box, push the button '**load tps file (no ruler/ no scale factor).**' Select the file you saved from TPSDIG. If you had a scale you could use the light blue boxes to set it, but we don't. A cloud of points will appear on the graph. This shows the position of each of the landmarks for each specimen as they have been superimposed. You are currently looking at a Bookstein coordinate superimposition.

The problem with two-point registration is that it takes all the variance from the two points that make up your baseline and redistributes it over the other points, because it holds those two points as stable all the change in position happens at the other points. This is not realistic, as change really happens throughout an organism, and is not limited to changing relative to two arbitrary points.

Procrustes deals with this problem by rescaling everything around the centroid of the object. The centroid is a point in the middle of the object calculated as the average of all the other points. First all the objects are moved, so that the centroid is located at the origin, thus eliminating location and two coordinates along with it. Next all the objects are rescaled so that their centroid size is one, eliminating size and one more dimension. Finally the objects are rotated, so that the **procrustes distances** between them are minimized, eliminating orientation and one more dimension. The **procrustes distance** is basically the distance between the objects in the new space defined by the first two transformations. I want to emphasize that the point of doing all this is to remove all the factors other than shape, not to reduce the number of dimensions. The dimensions are reduced as a byproduct of these transformations.

Go to the **green box** and toggle between the **show bc** button and **show procrustes** buttons. What differences do you see? Do you know why? Can you identify where your baseline is in the bc?

We could save the results of these superimpositions, in the **lavender box** you can see the buttons used for saving the results. The saved results files could be read by the other programs we'll be using. However, there is no need, because we'll be using other files that have many more samples for the remaining exercises. Our current example