G562 Geometric Morphometrics

Basic statistical analysis workflows

PAST

- Cite PAST as Hammer, Ø., Harper, D.A.T., Ryan, P.D. 2001. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica 4(1): 9pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm
- Find complete PAST manual at the software website: https://folk.uio.no/ohammer/past/faq.html

Preliminaries

- Load landmarks using "Open" in the Past3 menu
- Select all of the data, then Procrustes superimpose using Transform > Landmarks > Procrustes (2D+3D) [note that the Keep Size and Add Size Columns will only work if you added scale to your photographs when you collected the landmarks]
- Do PCA to obtain scores by choosing Geometry > Landmarks (2D) or Landmarks (3D) > PCA (relative warps)
- In the results window, choose the Scores tab and copy these to the clipboard using the button at the bottom of the window
- Paste the scores into a blank Excel spreadsheet and save in comma delimited format (.CSV)
- Note the number of PCs that PAST produces. It will either be 2k-4 (where k is the number of landmarks) or n-1 (where n is the number of specimens), whichever is smaller. This number is equal to the true degrees of freedom. Other programs may return 2k PCs no matter what, but the numbers in the extra columns will essentially be zeros with a bit of rounding error because there is no real shape variance associated with them.
- You can use Excel to add columns that contain other variables (e.g., body mass, habitat category, etc.) or you can add them in PAST itself after you import the scores.
- Open the spreadsheet in PAST. Your data should have both row and column labels so make sure that the Names,data option is ticked for both rows and columns on the next dialog box. If you saved as CSV then the separator will be Comma. Finally click the Import button at the bottom of the dialog box
- In PAST you can add variables for analysis as follows:
  - highlight first column (PC1)

- Edit > Insert More Columns
- To add a title for the new columns, click "column attributes" and change the name from its default value of c1.
- If you are adding a continuous variable, then unclick column attributes.
- If you are adding a categorical (group) variable, choose Group from the pulldown menu on the Type line before you unclick column attributes.
- Now enter data into the appropriate cells in the new column. You can copy and paste them from another spreadsheet or type them in

MANOVA (Multivariate Analysis of Variance)

MANOVA tests for differences in shape between groups.

To perform a MANOVA test you need a variable with group names (these can be letters like A, B, C, etc. or other text labels like Forest, Grassland, etc. That variable must be set to Type=Group by highlighting the column, clicking the "Column attributes button" and selecting Group from the pulldown list. A blue G should appear next to the column name once you've done this.

Terminology: PAST uses the terms "independent" and "dependent" for the kinds of variables in the tests. For geometric morphometrics, the dependent variables are always shape variables (either PC scores or superimposed landmark coordinates).

- Select the grouping variable column and the PC columns. If you have many data variables, you will need to move the grouping variable so that it is positioned just before the PC1 column. To do this, click "Drag rows/columns" from the top of the window, drag the column into the correct position, and click on "Select" to return to the regular mode.
- Ideally you want to select the grouping variable and **all** of the PC columns, but if you have more landmarks than specimens PAST may complain that there are too many variables. If so, select all of the PCs except the highest one. If PAST still complains, remove the next highest PC until the test runs.
- Perform the MANOVA test using Multivariate > Tests > One-way PERMANOVA. This version of MANOVA uses a permutation test to determine the significance. You can, in principle, use the ordinary MANOVA test from the pull-down menu, but it is a parametric test that requires the assumptions on this page: https://en.wikiversity.org/wiki/Analysis_of_variance/Assumptions . Permutation tests do not require assumptions like normality of data that parametric versions of MANOVA do.
- The permutation MANOVA should be run on geometric morphometric data using the "Euclidean" similarity index. The Summary page gives a p-value for the full multivariate test. It tells you whether any one or more of the groups is significantly different from the others. The Pairwise tab gives you a p-value whether individual groups are different from the others. These values are what are known as *post hoc* tests. Normally one adjusts the p-value for the fact that many tests are being performed, which increases the probability of a

false-positive result. Choose Bonferroni-corrected p-values from the pulldown tab (note that if you only have two groups, the corrected p-value is identical to the uncorrected value).

Multivariate linear regression

Regression is used to test the relationship between shape and a continuous variable like body mass. To perform a regression you need to add an independent continuous variable to your data set. You can change the default name of the variable in PAST just like with MANOVA, but do not reset its type (the default "—" should be used for a continuous variable).

Terminology: PAST uses the terms "independent" and "dependent" for the kinds of variables in the tests. For geometric morphometrics, the dependent variables are always shape variables (either PC scores or superimposed landmark coordinates).

- Select the grouping variable column and the PC columns. If you have many data variables, you will need to move the grouping variable so that it is positioned just before the PC1 column. To do this, click "Drag rows/columns" from the top of the window, drag the column into the correct position, and click on "Select" to return to the regular mode.
- PAST offers two options for geometric morphometric regression:
    - Option 1: Geometry > Landmarks (2D) > Multivariate linear regression (1 independent, n dependent)
        - This method is performed on the Procrustes superimposed landmarks, not on the PC scores. It allows you to visualize the relationship between shape and your independent variable using thin plate spline deformation grids.
        - Add continuous variable to the matrix of Procrustes superimposed landmarks as described above. Perform the test.
        - PAST may complain that there aren't enough observations. In this case you should still select all the landmark coordinates or else the entire shape won't be included in the regression. You can ignore the error message.
        - The statistics tab gives you the p-value for the overall multivariate test. Even though the results are presented in what is called a "MANOVA table", this is a regression.
        - The plot tab allows you to create graphs of one landmark coordinate at a time regressed onto the continuous variable. These graphs aren't particularly useful in most cases. The tab also give you univariate results with the intercepts and p-values. These results are also not particularly useful. Note that you can have a significant multivariate result (statistics tab) without any

        of the univariate results being significant or vice versa. Use the results of the multivariate results for your interpretation whether the relationship is significant.

        - The Numbers tab gives you all the univariate results in a table.
        - The Deformations tab allows you to plot shape models that correspond to your continuous variable. The continuous variable will show up as a named slider in the lower part of the right hand column of the tab. For example, as you change the value in the slider, the landmark grid changes shape to match what is predicted by the regression for that value of the continuous variable.
    - Option 2: Model > Linear > Multivariate (1 independent, n dependent)
        - This method is performed on the PC scores, just like the MANOVA test. You cannot visualize the results as thin plate spline grids if you use this method, but it should produce identical statistical results as Option 1.
        - Add continuous variable to the matrix of PC scores as described above. Perform the test.
        - Ideally you want to select the grouping variable and **all** of the PC columns, but if you have more landmarks than specimens PAST may complain that there are too many variables. If so, select all of the PCs except the highest one. If PAST still complains, remove the next highest PC until the test runs without an error.
        - The statistics tab gives you the p-value for the overall multivariate test. Even though the results are presented in what is called a “MANOVA table”, this is a regression.
        - The Plot tab allows you to plot a PC of your choice against the continuous variable to see what the slope looks like in this dimension. Remember that the regression has as many dimensions as there are PCs. This tab also gives univariate results. While you should always use the multivariate p-value from the statistics tab to develop your overall conclusion whether shape is related to the continuous variable, the results here may be of interest because some individual PCs may have a stronger relationship than others.
        - The Numbers tab gives the univariate results in a table form.
        - Note that you can your data may have a significant result in the overall multivariate test (Statistics tab) without any of the univariate results being significant, or vice versa. This is normal. Use the overall result to draw your conclusions.

Discriminant function analysis (DFA) / Canonical Variate Analysis (CVA)

DFA is an ordination method used to find the shape that separates two groups. It can be used to classify unknown specimens into a group based on a training set where the group membership is known.

To perform a DFA you need a variable with group names (these can be letters like A, B, C, etc. or other text labels like Forest, Grassland, etc. That variable must be set to Type=Group by highlighting the column, clicking the "Column attributes button" and selecting Group from the pulldown list. A blue G should appear next to the column name once you've done this.

The group value needs to be filled out for many specimens, but some can be left blank if the group is unknown. The algorithm will find a fit based on the known sample and use it to classify the unknowns.

- Select the grouping variable column and several of the PC columns. If you have many data variables, you will need to move the grouping variable so that it is positioned just before the PC1 column. To do this, click "Drag rows/columns" from the top of the window, drag the column into the correct position, and click on "Select" to return to the regular mode.
- DFA is not a statistical test and here you should <u>not</u> select all of the PCs. Doing so tends to overinflate the confidence of assigning groups and often results in unknown individuals not being classified.
- Perform the test with Multivariate > Ordination > Discriminant Analysis (LDA) [note that LDA stands for linear discriminant analysis, which is simply another name for DFA). Because this is not a statistical test, there are no p-values.
- The Plot tab shows the DFA plot. To understand it, turn on Convex Hulls and Group Labels. It will show how well the groups can be separated. There are as many axes as there are groups. Typically the first axis separates the first two groups, the second axis separates a third group, the third axis will separate a fourth group, and so on.
- The scores tab gives you the DFA scores in case you want to plot them in other software.
- The loadings tell you how much each PC contributes to the separation between groups. Large absolute values mean that a PC contributes heavily, values near zero mean it does not.
- The Classifier tab shows the real group ("given group"), the group that DFA estimates for each specimen along with a second "jackknifed" estimate, which is usually the more realistic estimate. The difference between the first classification and the jackknifed is that in the first one the specimen itself was used to make the separation so the classification is a bit circular. In the jackknifed classification the specimen being classified was omitted from the separation step so it is more fair (this is sometimes known as "leave-one-out cross validation"). If you had any unknown groups, you will see the classification given by DFA in the jackknifed column.

- The Confusion Matrix tab is a guide to determining how reliable the classifications are. The rows are the real group identity and the columns are the predicted classification. If the method works perfectly the sums of the rows and the columns will be equal. On the right it gives the percentage of observation that were correct in case you find the Confusion Matrix confusing, as many people do.

MorphoJ

- Cite MorphJ as C. P. Klingenberg. 2011. MorphoJ: an integrated software package for geometric morphometrics. Molecular Ecology Resources 11: 353-357.
- Full MorphJ documentation is here: http://www.flywings.org.uk/MorphoJ_guide/frameset.htm?index.htm

Preliminaries

- Start a new project in the File Menu.
- MorphoJ is picky about the TPS file format: make sure you changed commas and tabs to spaces or else it won't load
- Once loaded, your landmarks appear under "newDataset" (or whatever you named it) in the Project Tree tab. Select them by clicking on the name.
- Procrustes superimposition by Preliminaries > New Procrustes Fit
    - Choose options Align by principal axes unless you want to try to specify an orientation for them based on the farthest right and left landmark point. This won't change the superimposition, but it will affect what orientation the points end up in after they've been superimposed.
    - Click "Perform Procrustes Fit"
    - The window that opens in the Graphics tab shows the mean shape in bold points with numbers and the individual shapes as clouds of landmarks around the mean
- Generate a Covariance matrix with Preliminaries > Generate Covariance matrices
    - Click on your landmarks and then click the generate button
    - The new covariance matrix and Procrustes coordinates will appear in the Project Tree
-

MANOVA

MANOVA tests for differences in shape between groups.

- You will need to create and import a file with "Classifier variables", which are the grouping categories. See instructions in the MorphoJ documentation under File Menu > Import Classifier Variables at http://www.flywings.org.uk/MorphoJ_guide/
- Select the landmarks under Project Tree ("newDataset" or whatever name you gave the landmarks when you created the project)
- Perform MANOVA by choosing Variation > Procrustes ANOVA. This performs a permutation-based MANOVA.
  - Choose landmarks in the Dataset line
  - Choose the variable of interest under the Individual line
  - Execute
- Results appear in the Results tab
  - Ignore the Centroid size results. Unless you explicitly scaled your landmark coordinates, this test will be meaningless.
  - Assess significance using the p-value under Shape, Procrustes ANOVA

Multivariate Regression

Regression is used to test the relationship between shape and a continuous variable like body mass.

- To perform a regression you need to create a file with "Covariates", which are continuous variables. See instructions in the MorphoJ documentation under File Menu > Import Covariates at http://www.flywings.org.uk/MorphoJ_guide/
- Select the landmarks under Project Tree ("newDataset" or whatever name you gave the landmarks when you created the project)
- Perform regression by choosing Covariation > Regression. This performs a permutation-based multivariate regression.
  - Choose landmarks in the Datasets window of the Dependent Variables column, followed by Procrustes coordinates underneath it, and then Variables Procustes coordinates under that
  - In the Independent Variables column select landmarks in the top Datasets window, Covariates in the Data matrices window, and the variable of interest in the variables window.
  - Choose permutation test option
  - Click Excute
- Results appear in several places:
  - Results tab shows the statistical output. The % predicted value is equivalent to R^2 (the amount of shape variation explained by the continuous variable) and P-value is the significance of the multivariate regression
  - Graphics > Regression > Shape changes tab shows the shape model for what part of shape is related to the continuous variable. By default it shows a vector diagram where the tails on the

landmarks show how they are deformed as your variable gets larger (lollipop graph). Right click to select a thin plate spline grid transformation diagram instead. If the lines are too short to see or the grid is barely deformed you can exaggerate the effect using the Set Scale factor option when you right click on the graph.

  - Graphics > Regression > Scores shows a typical-seeming bivariate plot of the relationship between shape (Regression score 1) and Body Mass. The Y axis is a composite shape value that attempts to summarize the full multivariate shape variation. Right click to turn on labels and confidence ellipses to get a better feel for the relationship between shape and your continuous variable.

geomorph (R package)

- Cite the software with version and the publication. Type in R console
  > citation(package="geomorph")
  Adams, D.C., E. Otarola-Castillo and E. Sherratt. 2014 geomorph: Software for geometric morphometric analyses. R package version XX.X. http://cran.r-project.org/web/ packages/geomorph/index.html
  Adams, D.C., and E. Otarola-Castillo. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. Methods in Ecology and Evolution. 4:393-399.
- Some geomorph documentation is here:
  http://people.tamu.edu/~alawing/materials/ESSM689/Quick_Guide_to_Geomorph_v2.0.pdf
  https://cran.r-project.org/web/packages/geomorph/geomorph.pdf
- The first of these guides gives better instructions than I can. The workflow is this:
  - Import tps files with *readland.tps* (see Section 1.1)
  - Procrustes superimposition with *gpagen* (see Section 3.1)
  - Plot PCA morphospace with *plotTangentSpace* (see Section 4.1)
  - Perform MANOVA or multivariate regression with *procD.lm* (see Section 3.3)

TPS series programs from Jim Rohlf

The TPS series of programs are extremely easy to use and run in Windows operating systems. See the general workflows handout for instructions on using *tpsRelw* to do PCA morphospace plots.

Regression is used to test the relationship between shape and a continuous variable like body mass. MANOVA tests for differences in shape between groups. The *tpsRegr* package does both of them. The help file or Readme.txt file explain the procedure.