

A Practical Introduction to Geometric Morphometrics for Archaeological Science

Dr. Christian Steven Hoggard

Introduction

This guide provides a “hands-on” step-by-step introduction for the application of geometric morphometric (GMM henceforth) methodologies in archaeological science (as conducted the R Environment). Using a published dataset this guide will guide the reader through four key GMM procedures: 1) data importing (and creation), 2) data transformation, 3) data analysis and 4) data visualisation. A Japanese translation of this documentation is also provided.

This practical constitutes the second workshop of the #StayHomeButStudy event, organised by Dr. Atsushi Noguchi, Dr. Christian Steven Hoggard and Prof. Ben Marwick and is tailored for Japanese archaeologists, researchers and enthusiasts.

About the Code, Packages and Data

The data used throughout this guide originates from Ivanovaitè et al (2020): “*All these Fantastic Cultures? Research History and Regionalization in the Late Palaeolithic Tanged Point Cultures of Eastern Europe*”, published in the European Journal of Archaeology (<https://doi.org/10.1017/eea.2019.59>). The data can be found on GitHub (<https://github.com/CSHoggard/-Eastern-Europe-Tanged-Points>) and the Open Science Framework (<https://osf.io/agrwb/>).

All code, data and markdown document (in HTML and PDF format) for this practical can be found on the workshop repository (https://github.com/CSHoggard/-japanworkshop2020tree/master/workshop_2).

The GMM procedure detailed below is grounded on two-dimensional outline analysis. In conducting outline analysis for this practical the follow packages (including their imported packages) are necessary:

- * **Momocs** (Version 1.2.9) <https://cran.r-project.org/web/packages/Momocs/index.html>
- * **tidyverse** (Version 1.3.0) <https://cran.r-project.org/web/packages/tidyverse/index.html>

For literature pertaining to outline analysis see:

- * Claude, J. (2008). *Morphometrics with R*. Springer Publishing.
- * Bonhomme, V., Picq, S., Gaucherel, C., & Claude, J. (2014). Momocs: Outline analysis using R. *Journal of Statistical Software*, 56: 1–24.
- * Caple, J., Byrd, J., & Stephan, C. N. (2017). Elliptical Fourier analysis: Fundamentals, applications, and value for forensic anthropology. *International Journal of Legal Medicine*, 131 (6): 1675–1690.
- * Ferson, S., Rohlf, F. J., & Koehn, R. K. (1985). Measuring shape variation of two-dimensional outlines. *Systematic Zoology*, 34 (1): 59–68.
- * Kuhl, F. P., & Giardina, C. R. (1982). Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing*, 18 (3): 236–258.
- * Yoshioka, Y. (2004). Analysis of petal shape variation of *Primula sieboldii* by elliptic fourier descriptors and principal component analysis. *Annals of Botany*, 94 (5), 657–664.
- * Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, C-21 (3): 269–281.

Software Installation

Following the installation of R or RStudio, we can now install the required packages:

```
if(!require("Momocs")) install.packages('Momocs', repos='http://cran.us.r-project.org')
if(!require("tidyverse")) install.packages('tidyverse', repos='http://cran.us.r-project.org')
if(!require("rio")) install.packages('rio', repos='http://cran.us.r-project.org')
```

To download the data into R/Rstudio we can use the `import` function from the `rio` package, and extract the data from the GitHub repository :

```
database <- rio::import("https://github.com/CSHoggard/-workshopjapan2020/raw/master/workshop_2/database")
tpsdata <- rio::import("https://github.com/CSHoggard/-workshopjapan2020/raw/master/workshop_2/tpslines")
```

The tidyverse and Momocs packages may take time to install given the size of the file, so please be patient. Once installed, the packages can be activated through the `library()` function:

```
library(Momocs)
library(tidyverse)
```

About the Data

This data was composed to assess the robustness of cultural taxonomies in the Final Palaeolithic period of Eastern Europe, as denoted through tanged point variants. It consists of 250 tanged point outlines and produced in the TPS Suite (<https://life.bio.sunysb.edu/morph/soft-dataacq.html>), using the *outline object* function. As this dataset was produced in the TPS Suite the file format is *.tps*. In their composition these outlines are semilandmarks, an algorithm-produced series of equidistant points are each shape. A database of all examples and their respective cultural assignment is also provided.

Importing Geometric Morphometric Data: Alternative Approaches

There are a number of ways which landmark and outline morphometric data can be imported into R. Here, for ease and replicability, the data was stored on a GitHub repository and directed into the R environment, utilising the `Momocs::import_tps()` function in a *.rds* file. Other ways to import *.tps* data (if saved locally) include the `geomorph::readland.tps()` and rewriting tools in Momocs e.g. `Momocs::rw_rule()`. Data from stereomorph can also be imported through the `Momocs::import_StereoMorph_ldk()` and `Momocs::import_StereoMorph_curve()` functions.

Within Momocs, outlines can be extracted from silhouette data through the `Momocs::import_jpg()` and `Momocs::import_jpg1()` functions. See their helpfiles for more details.

Examining the Data

We can now preview our database using the `head()` function, and call our `tpsdata` through the `View` functions:

```
head(database)
```

```
## # A tibble: 6 x 10
##   ID Site Context Longitude Latitude Country Archaeological_~ File_Name
##   <dbl> <chr> <chr>      <dbl>      <dbl> <chr>    <chr>              <chr>
## 1     1 Balt~ Baltaš~    24.0      54.0 Lithua~ Baltic Magdalen~ Bal.1
## 2     2 Baro~ Barouka    30.3      53.5 Belarus Grensk        Bor.1
## 3     3 Baro~ Barouka    30.3      53.5 Belarus Grensk        Bor.3
## 4     4 Baro~ Barouka    30.3      53.5 Belarus Grensk        Bor.4
## 5     5 Baro~ Barouka    30.3      53.5 Belarus Grensk        Bor.5
## 6     6 Baro~ Barouka    30.3      53.5 Belarus Grensk        Bor.6
## # ... with 2 more variables: Reference <chr>, Notes <chr>
```

We can observe that the group data (Archaeological_Unit) is of type 'character' and not 'factor', as required for our analysis. This can be corrected through the `base::as_factor()` function:

```
database$Archaeological_Unit <- as.factor(database$Archaeological_Unit)

is.factor(database$Archaeological_Unit) # check to see the data is now of type 'character'
```

```
## [1] TRUE
```

We can now inspect the number of different archaeological units within our dataset through the `base::summary()` function:

```
summary(database$Archaeological_Unit)

##      Baltic Magdalenian Bromme (Eastern Europe) Bromme (Western Europe)
##              36                      9                      49
##      Grensk              Krasnosillya              Perstunian
##              55                      29                      4
##      Pitted Ware (Type A)              Podolian              Vyshegorian
##              24                      14                      8
##      Wolkushian
##              22
```

The `base::View()` function will highlight the three constituent parts of the tps file: the 1) *Coo* (coordinate data), 2) *cur* (the curve data if present), and 3) *scale* (the scale data if present). It is the *Coo* data which we will take forward, with the database, to examine shape variation among our tanged points.

GMM Procedure 1: Outline file creation

Central to Momocs are a specific suite of shape classes: 1) *outlines* (OutCoo), *open outlines* (OpnCoo) or *landmarks* (LdkCoo). While some operations in Momocs are generic and do not depend on one of these classes, many functions require your data to be one of these S3 objects. In this instance our data is comprised of outlines, and so we wish for our data to be OutCoo to enable *efourier* (elliptic Fourier), *rfourier* (radii Fourier) and *tfourier* (tangent angle Fourier) methodologies. The coordinate data (*coo*) must therefore be turned into outline data through the `Out()` function:

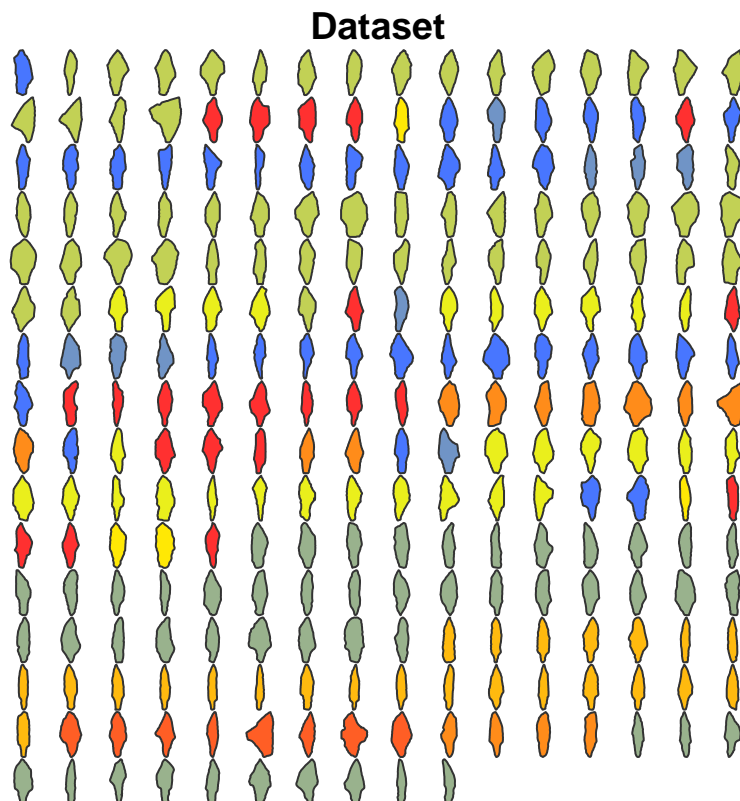
```
shape <- Out(tpsdata$coo, fac = database) # incorporating our database as our factors
shape # call the object
```

```
## Out (outlines)
## - 250 outlines, 1543 +/- 1370 coords (in $coo)
## - 10 classifiers (in $fac):
## # A tibble: 250 x 10
##   ID Site Context Longitude Latitude Country Archaeological_~ File_Name
##   <dbl> <chr> <chr>      <dbl>      <dbl> <chr>      <fct>      <chr>
## 1     1 Balt~ Baltaš~    24.0      54.0 Lithua~ Baltic Magdalen~ Bal.1
## 2     2 Baro~ Barouka    30.3      53.5 Belarus Grensk      Bor.1
## 3     3 Baro~ Barouka    30.3      53.5 Belarus Grensk      Bor.3
## 4     4 Baro~ Barouka    30.3      53.5 Belarus Grensk      Bor.4
## 5     5 Baro~ Barouka    30.3      53.5 Belarus Grensk      Bor.5
## 6     6 Baro~ Barouka    30.3      53.5 Belarus Grensk      Bor.6
## # ... with 244 more rows, and 2 more variables: Reference <chr>, Notes <chr>
## - also: $ldk
```

GMM Procedure 2: Outline Visualisation

Now our data is in the R environment, and in the appropriate class, we can examine the outlines. We can first look at all outlines through the `Panel()` function. Factors can also be coloured in using the `fac` argument in `Panel()`.

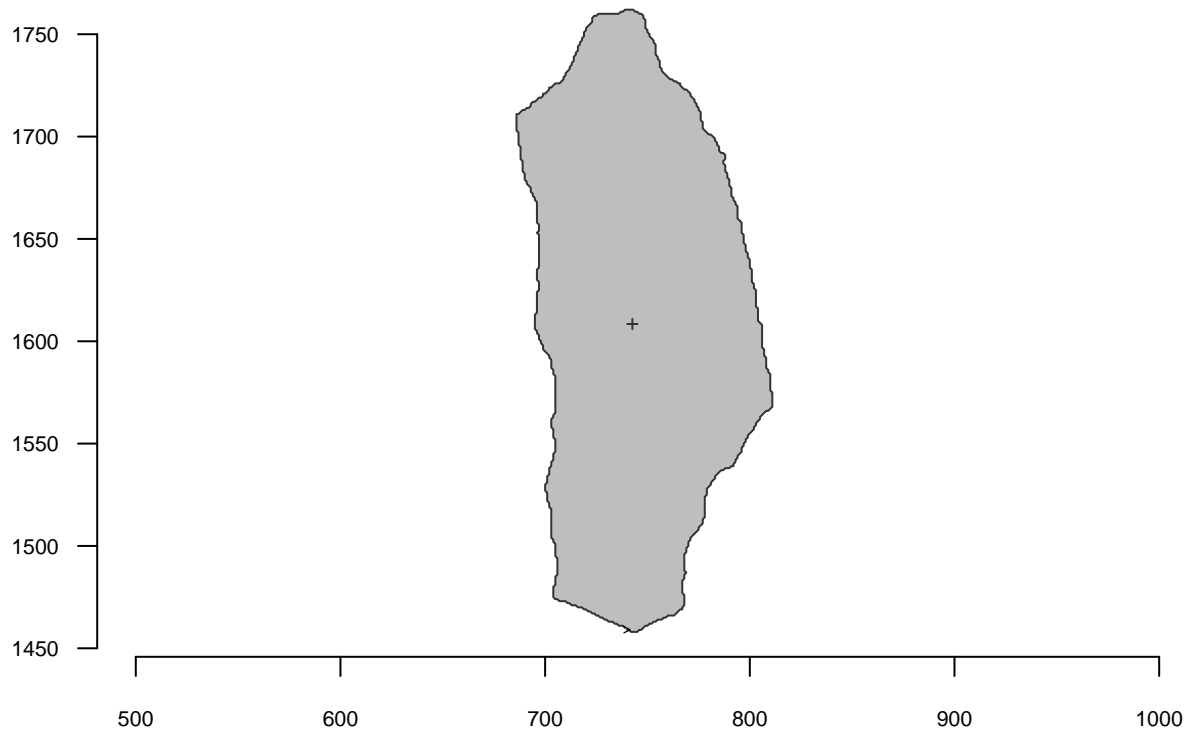
```
panel(shape, fac = 'Archaeological_Unit', main = "Dataset")
```



XXXXXXXXXXXXXXXXXXXX

```
coo_plot(shape[1], col = "grey", main = "Database Artefact #1")
```

Database Artefact #1

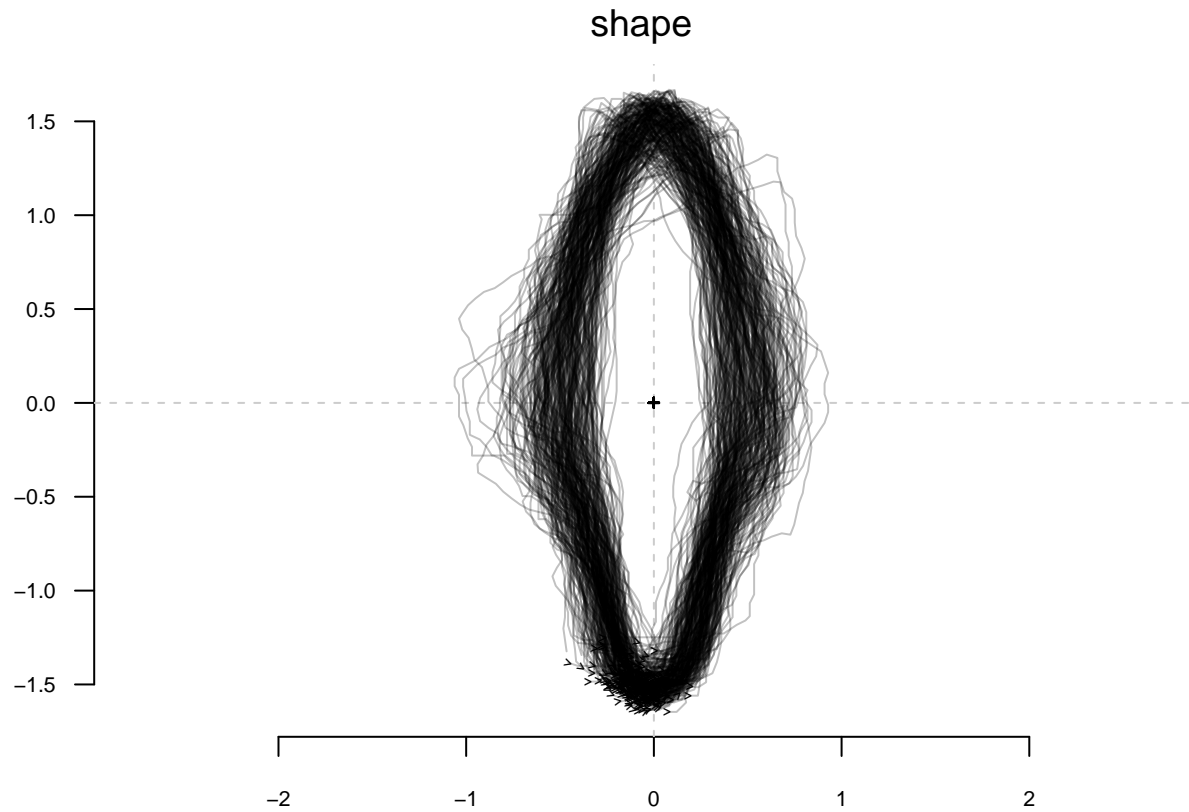


XXXXXXXXXXXXXXXXXXXX

GMM Procedure 3: Outline Normalisation

```
shape <- coo_center(shape)
shape <- coo_scale(shape)
shape <- coo_close(shape)

stack(shape, main = "")
```

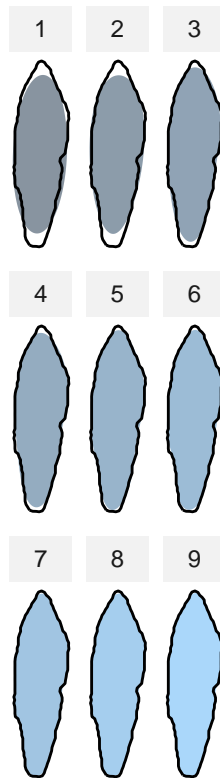


GMM Procedure 4: Elliptic Fourier Transformation

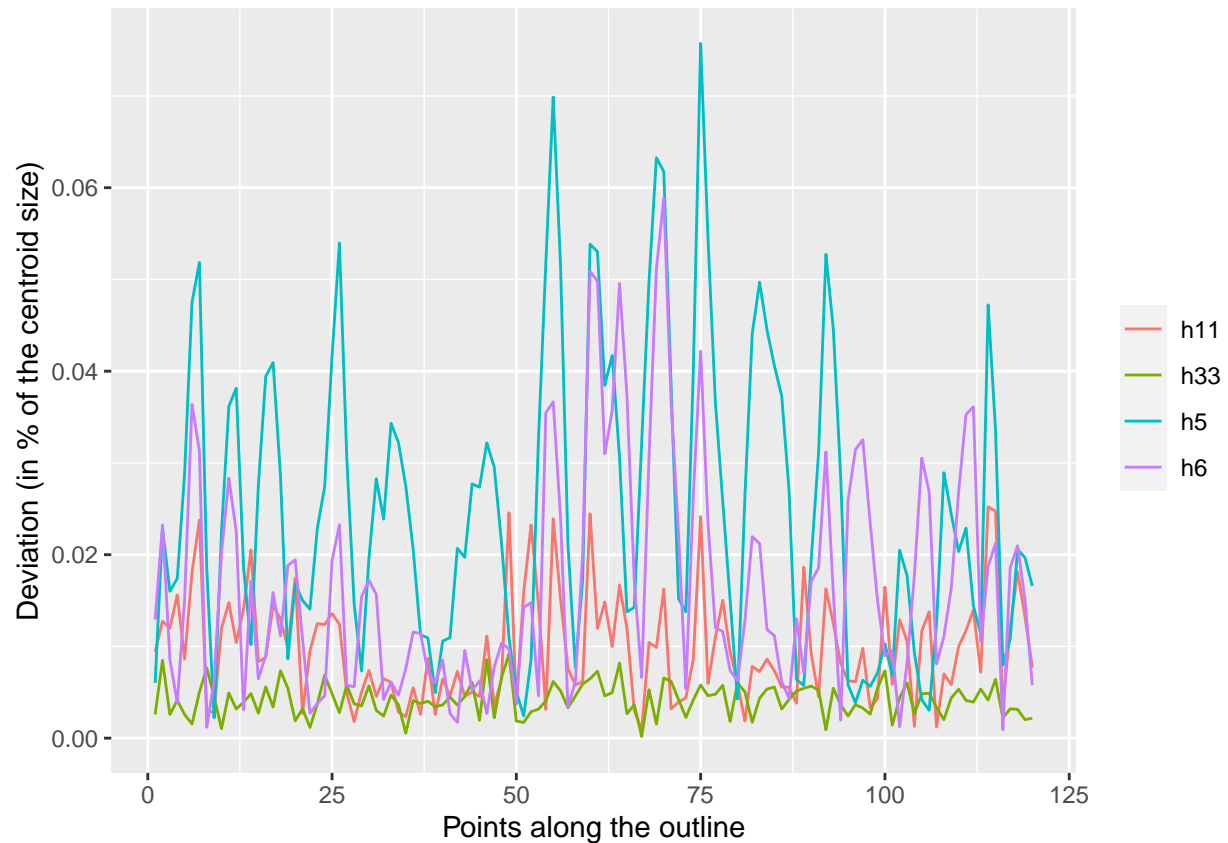
XXXXXXXXXXXXXXXXXXXX

```
calibrate_reconstructions_efourier(shape)
```

50) Chvojnjaja



```
calibrate_deviations_efourier(shape)
```



XXXXXXXXXXXXXXXXXXXX

```
efashape <- efourier(shape, nb.h = 11, smooth.it = 0, norm = TRUE)
efashape
```

```
## An OutCoe object [ elliptical Fourier analysis ]
## -----
## - $coe: 250 outlines described, 11 harmonics
## # A tibble: 250 x 10
##   ID Site Context Longitude Latitude Country Archaeological_~ File_Name
##   <dbl> <chr> <chr>      <dbl>    <dbl> <chr>    <fct>          <chr>
## 1     1 Balt~ Baltaš~    24.0    54.0 Lithua~ Baltic Magdalen~ Bal.1
## 2     2 Baro~ Barouka    30.3    53.5 Belarus Grensk        Bor.1
## 3     3 Baro~ Barouka    30.3    53.5 Belarus Grensk        Bor.3
## 4     4 Baro~ Barouka    30.3    53.5 Belarus Grensk        Bor.4
## 5     5 Baro~ Barouka    30.3    53.5 Belarus Grensk        Bor.5
## 6     6 Baro~ Barouka    30.3    53.5 Belarus Grensk        Bor.6
## # ... with 244 more rows, and 2 more variables: Reference <chr>, Notes <chr>
```

GMM Procedure 5: Exploratory Procedure #1: PCA

XXXXXXXXXXXXXXXXXXXX

```
pcashape <- PCA(efashape)
PCcontrib(pcashape)
```

```
## `prop` not provided. All axes returned
```