

Momocs : outline analysis using R

Vincent Bonhomme, Sandrine Picq, Cedric Gaucherel, Julien Claude

► To cite this version:

Vincent Bonhomme, Sandrine Picq, Cedric Gaucherel, Julien Claude. Momocs : outline analysis using R. Journal of Statistical Software, University of California, Los Angeles, 2014, 56 (13), 24 p. 10.18637/jss.v056.i13 . hal-02635896

HAL Id: hal-02635896

<https://hal.inrae.fr/hal-02635896>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License



Journal of Statistical Software

February 2014, Volume 56, Issue 13.

<http://www.jstatsoft.org/>

Momocs: Outline Analysis Using R

Vincent Bonhomme
French Institute of Pondicherry

Sandrine Picq
UMR CBAE, Montpellier

Cédric Gaucherel
French Institute of Pondicherry

Julien Claude
University of Montpellier II

Abstract

We introduce here **Momocs**, a package intended to ease and popularize *modern morphometrics* with R, and particularly outline analysis, which aims to extract quantitative variables from shapes. It mostly hinges on the functions published in the book entitled *Modern Morphometrics Using R* by Claude (2008). From outline extraction from raw data to multivariate analysis, **Momocs** provides an integrated and convenient toolkit to students and researchers who are, or may become, interested in describing the shape and its variation. The methods implemented so far in **Momocs** are introduced through a simplistic case study that aims to test if two sets of bottles have different shapes.

Keywords: modern morphometrics, comparison of shapes, Fourier transform, outline analysis, thin plate splines, R.

1. Introduction

1.1. The aim and purpose of morphometrics

The link, if there were one, between the form and the function of objects, living or inert, has been one of the most enduring questions in the realm of science. In many situations, analyzing the shape variation among objects can bring deep insights into their functioning and to the underlying mechanism leading to their variation in shape. For instance, an evolutionary biologist may be interested in testing the proximal link between the shape and the photosynthetic capacities, hydraulic properties, or the way leaves develop from a bud. Similarly, distal causes such as local adaptations in comparing different populations, or the diversification of foliar organs along the evolutionary lineage of a species, can be investigated. In a very different

field, geographers would like to test if differences in the shape of cities can be attributed to external factors such as highways crossing them or oceans bordering them. In addition, morphometrics, the so-called quantitative study of form, and derived techniques such as shape clustering and recognition, can concern a wide spectrum of disciplines.

Morphometrics aims at analyzing the variation and covariation of the *size* and *shape* of objects, defining altogether their form. Shape and form might be confusing words, used as synonyms in many languages. Hereafter, we will use the definition of shape proposed by Kendall (1989) and Small (1996) that it is “the total of all information invariant under translations, rotations, and isotropic rescaling”. What remains if we draw a heart, and then rotate the page, or change its size uniformly, or move the page about on the desk, is the shape of the heart. Shape by essence is better described in three dimensions, except for flat objects (see Kendall 1989, for a review). In this paper, however, only two dimensional shape will be considered, e.g., three-dimensional objects will be viewed from one side and considered as represented by their projections on a plane.

How to quantify the shape of the heart that was described above, and how to compare a given number of hearts to show variations among themselves? This paper aims to provide a didactic introduction to the shape concept, and particularly a branch of morphometrics relying on outline analysis. Finally, a case study of outline analysis is presented; it uses **Momocs**, the R package presented here, whose name stands for *modern morphometrics*, and is intended to simplify such morphometrical analyses.

1.2. How to compare shapes?

The everyday approach to describing shapes is to use words, such as “round”, “narrow”, “heart-shaped”, and “symmetric”. Such usage points to a very limited vocabulary access: shapes can be too complex, and differences between them too subtle for words. More importantly, they are *ad hoc* descriptors. In describing the shape of a different object, one will probably not use the same vocabulary thus making comparison hard to appreciate.

A quantitative framework was introduced by *traditional morphometrics* which measures distances, areas, etc. and compares them in an uni- or multivariate framework (see Rohlf and Marcus 1993). For instance, in describing and comparing several human faces, the lengths of the ears and the noses, the interpupillary distances, and other lengths or ratios of lengths, taken homogeneously between the individuals, are used to test differences between genders or the covariances between parts of the body. In his seminal book *On Growth and Form*, Thompson (1917) compared shapes in what could be called today a morphometrics approach, and offered new ways of understanding their variations: i) some changes in the developmental processes of living organisms, while minor, can lead to dramatic morphological changes, and ii) physical constraints such as growing mechanics are of first importance in the final form of organisms (Figure 1). Morphometrics contributed towards raising development from a simple bridge between genes and organisms to a central catalyst of evolutionary change between species. Many attempts have also been made outside biology, notably using scalar indices to describe and compare shapes: compactness, elongation, fractal dimension, etc. are such examples used to compare boundaries of political or physical geographical objects such as cities, states, or watersheds, etc. (Moellering and Rayner 1981; Wentz 2000; Miller and Wentz 2003). In particular **Momocs** and the R routines, on which it hinges, have already been used to study the influence of height and body mass on human bodies outlines (Courtiol, Ferdy,

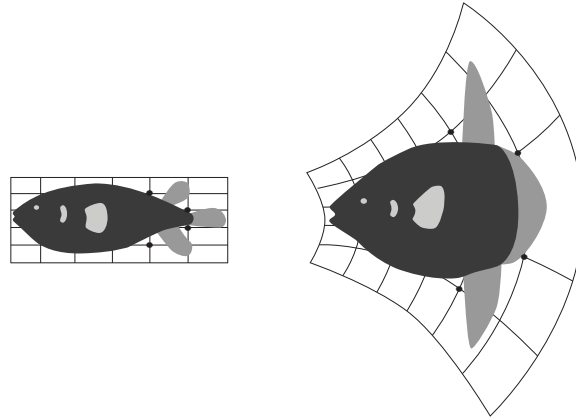


Figure 1: These popular D'Arcy Thompson's fishes illustrate his central thesis: "An organism is so complex a thing, and growth so complex a phenomenon, that for growth to be so uniform and constant in all the parts as to keep the whole shape unchanged would indeed be an unlikely and an unusual circumstance. Rates vary, proportions change, and the whole configuration alters accordingly" (cited from [Thompson 1917](#)). From a morphological point of view, these two fishes seem very different, but a little change in the growth rate of the caudal (right) part, here illustrated by the grid and four particular points taken on it, leads to dramatic change. This example serves to demonstrate that, by providing a rigorous method of describing shape, morphometrics can help a better understanding of how entities in nature function. Redrawn after [Thompson \(1917\)](#).

[Godelle, Raymond, and Claude 2010](#)), to provide complementary information in the shape description of watersheds where only scalar indices are classically used ([Bonhomme, Frelat, and Gaucherel 2013a](#)) and the first evidence of intraspecific variability in the shape of pollen grains of anemophilous species ([Bonhomme, Prasad, and Gaucherel 2013b](#)).

At the other end of the XXth century, computerized data acquisition and treatment arose synchronously with an array of new methodological developments and these techniques of *modern morphometrics* revolutionized the historical scope of morphometrics ([Rohlf and Slice 1990](#); [Rohlf and Marcus 1993](#); [Adams, Rohlf, and Slice 2004](#)).

1.3. Modern morphometrics: Landmark configuration and outline analysis

Modern morphometrics considers shape as a whole, taking into account all the geometrical relationships of the input data. The two main approaches in use are: the study of landmark configurations, and outline and surface analyses. Both of them preserve the geometrical information, i.e., relative positions between all points are kept ([Moellering and Rayner 1981](#); [Kuhl and Giardina 1982](#)). This allows shape reconstruction from their numerical signature, a fact which is of great interest since we can then define the most frequent shapes, the rare or the impossible ones, infer intermediate shapes, etc. in other words, these approaches can reveal some functional links between the shape and its variation with the processes leading to it.

Configuration of landmarks can be summarized as follows: the relative positions of a set of points, called landmarks, are considered globally e.g., by using a matrix of their pairwise

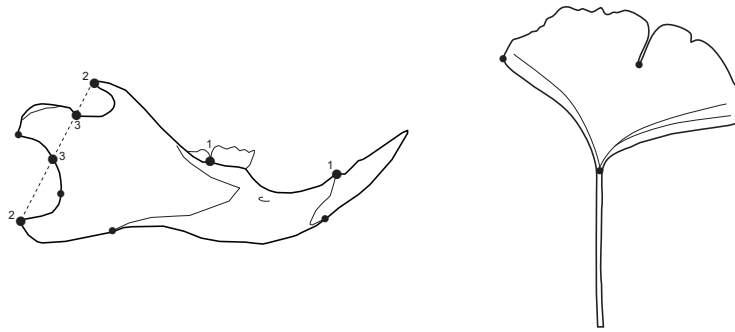


Figure 2: Two biological shapes that can be explored with modern morphometrics. The mouse jaw on the left, redrawn after [Claude \(2008\)](#), exhibits many potential landmarks. Some of them are (1) frontiers between bones, (2) sharp angles on a single bone, and (3) the region with high curvature. The same approach cannot be easily applied with the *Ginkgo biloba* leaf on the right since less landmarks can be identified or are not consistently present on *Ginkgo* leaves, for instance the separation between the two lobes. Outline analysis can become a solution in cases such as this.

euclidean distances ([Richtsmeier, Cheverud, and Lele 1992](#); [Richtsmeier, Burke Deleon, and Lele 2002](#)). The landmarks need to be structurally similar (ideally homologous) between individuals. For instance, if the variation of the vertebrate skull is considered, the bones that share a common origin are defined as homologous and can be part of a configuration of landmarks along with any other geometrical feature that can be unambiguously identified (Figure 2, see also [Macleod 1999](#), for a general discussion on homology).

Other objects, for instance a leaf, do not display any or display too few landmarks. Methods for analyzing configurations of landmarks cannot be easily applied here, and outline analysis will be used instead. This second approach considers outline as a whole. Outline is defined here as the closed polygon formed by the $(x; y)$ coordinates of pixels defining it. One popular outline analysis approach, on which **Momocs** is focused so far, uses Fourier series to describe shapes, and is detailed below. Finally, three dimensional surfaces or outlines are an other area of morphometrics, but will not be considered here.

1.4. Momocs: Analysis of outline variation using R

This paper introduces **Momocs** (<http://CRAN.R-project.org/package=Momocs>), a package to analyze outlines of shapes, using R ([R Core Team 2013](#)). The package originated from some core functions published by one of us ([Claude 2008](#)) and reviewed by [Bowman \(2009\)](#). These functions were turned into an integrated framework and a standalone R package. Below are provided step-by-step guidelines for performing *modern morphometrics*, from which **Momocs** derives its name, using R. The package's vignette *A Graphical Introduction to Momocs and Outline Analysis Using R* ([Bonhomme 2012](#)) also provides an extensive description of the functions of the package.

While both outline analysis and R (to a much greater extent) have been used in increasing measure, so far no dedicated tool has been available at CRAN. **Momocs** aims to fill this gap. Other tools exist, but they focus on configuration of landmarks: **shapes** by [Dryden \(2012\)](#), **MorphoJ** by [Klingenberg \(2011\)](#), and recently **geomorph** by [Adams and Otarola-Castillo](#)

(2012). Outline analyses can be performed by some standalone programs but often not under an open license and only on certain operating systems. The Stony Brook University webpage lists most of them (<http://life.bio.sunysb.edu/morph/>), the **SHAPE** suite by Iwata, Ne-sumi, Ninomiya, Takano, and Ukai (2002) being broadly used. **Momocs** is placed under the GPL license and through this license, it may become a collaborative hub for other researchers to explore new methods or implement existing approaches.

2. Mathematical background of Fourier-based outline analyses

Approaches for analyzing outlines rather than landmark configurations estimate parameters of functions rather than the relative positions of landmarks after superimposition. One of the strategies is to adjust Fourier series to some shape descriptors, and this is the main approach followed, so far, by **Momocs**.

2.1. Fourier transformations and closed outlines

Outline analysis does necessarily require that the outline on which points are sampled to be structurally defined: what is extracted is the geometrical information contained in the outline itself and taken as a whole. However, when comparing shapes, outlines should correspond to structurally similar features. In particular, Fourier-based approaches are powerful enough to extract this geometric information. They have their basis in the idea of Fourier series: to decompose a periodic function into a sum of more simple trigonometric functions such as sine and cosine. These simple functions have frequencies that are integer multiples, i.e., are harmonics, of one another. The lower harmonics provide approximations for the coarse-scale trends in the original periodic function, while the high-frequency harmonics fit its fine-scale variations.

Fourier series can be used in morphometrics, amongst many other derived applications since closed outlines can be considered as periodic functions. If we start somewhere on the outline and follow it, we will pass again and again by the same starting point and thus periodic functions can describe this outline. Such functions are: the distance of any point on the outline to the centroid of the shape, the variation of the tangent angle for any point, or the $(x; y)$ coordinates on the plane. A, or several, periodic functions are then obtained and can be decomposed (and thus described) by Fourier series. These three different methods are available in **Momocs**, and hereafter called “radius variation”, “tangent angle” and “elliptical analysis” (Figure 3), and their comparison has been discussed extensively by Rohlf and Archie (1984).

The principle of Fourier series described above applies to continuous functions. Since a shape is based on a finite number of discrete points, typically coordinates on a plane (or a space), a discrete equivalent of Fourier series is used in morphometrics. A given number of points called pseudo-landmarks, have to be sampled along the outline before performing analysis of outline variation. All Fourier decompositions result in an harmonic sum of trigonometric functions weighted with harmonic coefficients. They are (usually) normalized to remove homothetic, translational or rotational differences between shapes. Two or four coefficients, depending on the approach used, are obtained for each harmonic calculated and can then be considered as quantitative variables. Nyquist frequency precludes more harmonics than half the number of points fitted, which is thus their upper limit. The geometrical information contained in

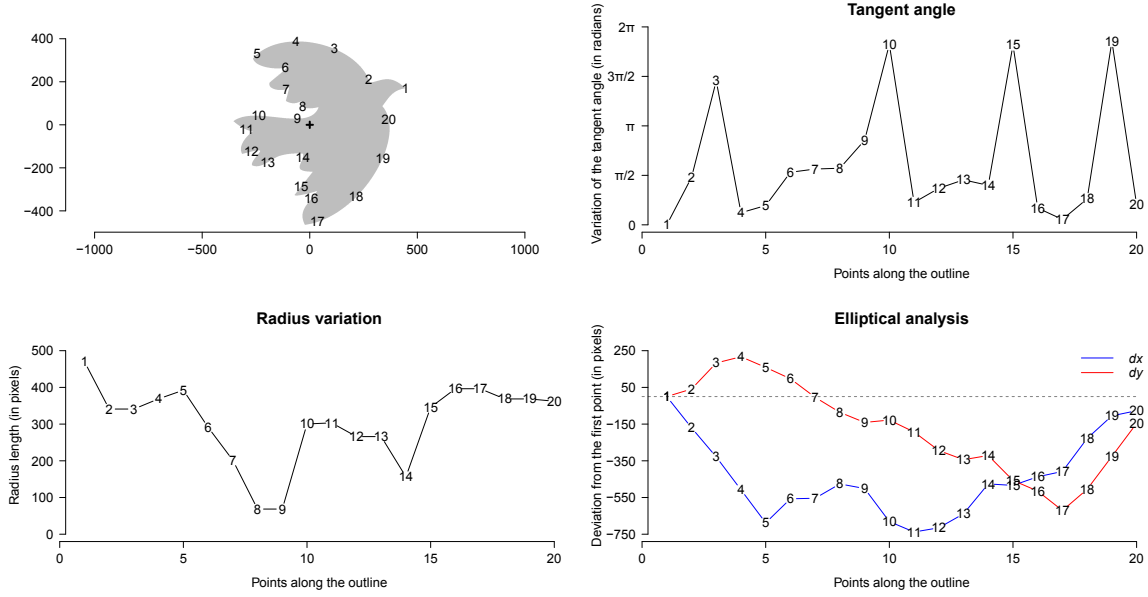


Figure 3: Twenty equidistant points have been sampled, starting from the beak, counter-clockwise and along the curvilinear abscissa of the dove shape (top left) outline, inspired by Picasso’s drawing. This outline can be described using Fourier-based methods. Tangent angle (top right) illustrates the variation of the tangent angle along the outline. Radius variation (bottom left) illustrates the length of the radius, here considered as the distance between the center of the shape (the cross within the dove outline) and the points along the outline. Elliptical analysis (bottom right) shows the two curves corresponding to $x_n - x_1$ (in blue) and $y_n - y_1$ (in red).

the outlines are thus quantified and can be analyzed with classical multivariate tools. The following sections detail the core functions of **Momocs**, and an extensive description can be found in [Rohlf and Archie \(1984\)](#) and [Claude \(2008\)](#).

2.2. Fourier radius variation

[Zahn and Roskies \(1972\)](#) stated that, given a closed outline, the radius r , taken as the distance from the outline centroid and a given point of the outline, can be expressed as a periodic function of the angle θ . Harmonics from 0 to k approximate the function:

$$r(\theta) = \frac{1}{2}a_0 + \sum_{n=1}^k a_n \cos(w_n \theta) + b_n \sin(w_n \theta)$$

with:

$$a_n = \frac{2}{p} \sum_{i=1}^p r_i \cos(n\theta_i), \quad b_n = \frac{2}{p} \sum_{i=1}^p r_i \sin(n\theta_i),$$

and:

$$a_0 = \sqrt{\frac{2}{p}} \sum_{i=1}^p r_i.$$

w refers to the pulse and p is the number of sampled points along the outline (equivalent here to the number of sampled radii in this case). The a_n and b_n harmonic coefficients, extracted for every individual shape, can be used for multivariate analyses to compare a set of outlines.

2.3. Fourier tangent angle

Radius variation may fail to fit some complex outlines, in particular when a given radius intercepts the outline twice, a situation that can arise when the outline presents convexities and concavities. [Zahn and Roskies \(1972\)](#) proposed also another approach. The Fourier tangent angle fits the cumulative change in the angle of a tangent vector ($\phi(t)$), as a function of the cumulative curvilinear distance t along the outline.

Given a closed outline, previously scaled to 2π , $\phi(t)$ can be expressed as:

$$\phi(t) = \theta(t) - \theta(0) - t,$$

where t is the distance along the outline, $\theta(t)$ the angle of the tangent vector at t and $\theta(0)$ the angle of the tangent vector taken for the first point. It can be removed for normalizing the coefficients obtained. Two coefficients per harmonic can be estimated as follows:

$$a_n = \frac{2}{p} \sum_{i=1}^p \phi(t) \cos(n\theta_i), \quad b_n = \frac{2}{p} \sum_{i=1}^p \phi(t) \sin(n\theta_i),$$

and:

$$a_0 = \sqrt{\frac{2}{p}} \sum_{i=1}^p \phi(t)$$

2.4. Elliptic Fourier analysis

The last approach presented here is due to [Giardina and Kuhl \(1977\)](#) and [Kuhl and Giardina \(1982\)](#) who developed a method for fitting separately the x and y coordinates of an outline projected on a plane. This method has become very popular since it has great advantages over the other Fourier-based approaches: equally spaced points are not required, virtually any outline can be fitted (see [Rohlf and Archie 1984](#); [Crampton 1995](#); [Renaud and Michaux 2003](#)) and the coefficients can be made independent of outline position and normalized for size.

Let T be the perimeter of a given closed outline, here considered as the period of the signal. One sets $\omega = 2\pi/T$ to be the pulse. Then, the curvilinear abscissa t varies from 0 to T . One can express $x(t)$ and $y(t)$ as follows:

$$x(t) = \frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t)$$

with:

$$a_n = \frac{2}{T} + \int_0^T x(t) \cos(n\omega t) dt, \quad b_n = \frac{2}{T} + \int_0^T x(t) \sin(n\omega t) dt.$$

Similarly,

$$y(t) = \frac{c_0}{2} + \sum_{n=1}^{+\infty} c_n(\cos n\omega t) + d_n(\sin n\omega t)$$

with:

$$c_n = \frac{2}{T} + \int_0^T y(t) \cos(n\omega t) dt, \quad d_n = \frac{2}{T} + \int_0^T y(t) \sin(n\omega t) dt.$$

Since the outline contains a finite number of points given by k , one can calculate discrete estimators for every harmonic coefficient of the n th rank:

$$a_n = \frac{T}{2\pi^2 n^2} \sum_{p=1}^k \frac{\Delta x_p}{\Delta t_p} \left(\cos \frac{2\pi n t_p}{T} - \cos \frac{2\pi n t_{p-1}}{T} \right),$$

$$b_n = \frac{T}{2\pi^2 n^2} \sum_{p=1}^k \frac{\Delta x_p}{\Delta t_p} \left(\sin \frac{2\pi n t_p}{T} - \sin \frac{2\pi n t_{p-1}}{T} \right).$$

$\Delta x_1 = x_1 - x_k$ and c_n and d_n are calculated similarly. a_0 and c_0 correspond to the estimate of the coordinates of the centroid of the original outline and are estimated by:

$$a_0 = \frac{2}{T} \sum_{i=1}^p x_i, \quad c_0 = \frac{2}{T} \sum_{i=1}^p y_i.$$

Intuitively, for all positive integers n , the sum of a cosine curve and a sine curve represent the n th harmonic content of the x and y projections of the k -edged polygon, and for any n , these two curves define an ellipse in the plane. [Ferson, Rohlf, and Koehn \(1985\)](#) noticed that in the “time”, say one period, it takes the n th harmonic to traverse its ellipse n times, the $(n + 1)$ th harmonic has traversed its own ellipse $n + 1$ times (Figure 4). [Ferson et al. \(1985\)](#) noticed that the reconstruction of the original polygon is done by vector-adding these ellipses for all harmonics, which echoes (the ancient astronomer) Ptolemy’s epicycles, and the reconstruction obtained from N harmonics is the best possible fit in a least-squares sense.

In elliptic Fourier analysis, four coefficients per harmonic are obtained, two for x and two for y . We can use the first harmonic, the one that defines the best-fitting ellipse, to normalize the harmonic coefficients and make them invariant to size and rotation. The harmonic coefficients can also be normalized for the location of the first outline coordinate. When this is done, the shapes are individually aligned according to their first fitted ellipse. If at least one homologous point can be defined, one can rather use them to align the outlines. Normalized elliptic Fourier coefficients, further symbolized by A_n , B_n , C_n and D_n , are obtained:

$$\begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix} = \frac{1}{\lambda} \begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix} \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \begin{pmatrix} \cos n\theta & -\sin \theta \\ \sin n\theta & \cos n\theta \end{pmatrix}$$

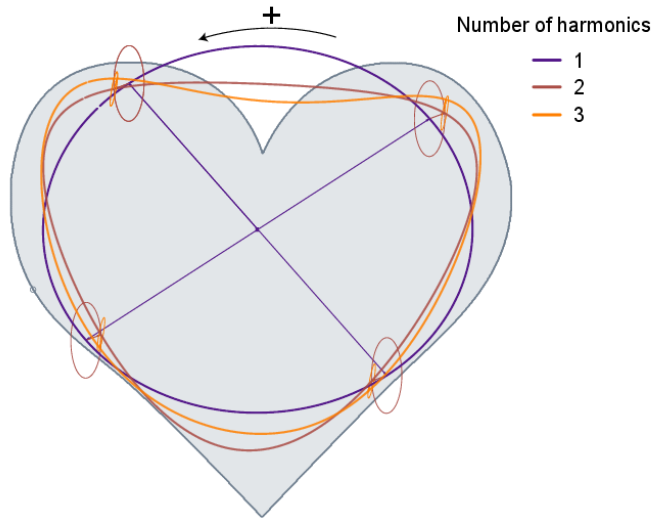


Figure 4: For all positive integers, the sum of a cosine curve and a sine curve defines an ellipse in the plane. Elliptic Fourier analysis is based on an harmonic sum of such ellipses as in Ptolemy’s astronomical system with higher harmonic order ellipses “rolling” within all lower order ellipses. Three harmonics are here shown at four locations on the original outline.

The scale λ is estimated as the magnitude of the semi-major axis of the ellipse as defined by the first harmonic. The second right term corresponds to the orientation of the first ellipse, with ψ being the rotation angle, the third to the original harmonic coefficient, and the last to the rotation of the starting point to the end of the ellipse, with a rotation angle of θ . [Ferson et al. \(1985\)](#) also supplied the following formulas with which to calculate these parameters:

$$\psi = 0.5 \arctan \frac{2 \times (a_1 b_1 + c_1 d_1)}{a_1^2 + c_1^2 - b_1^2 - d_1^2}$$

with first

$$\lambda = \sqrt{a^{*2} + c^{*2}}$$

and

$$\theta = \arctan(c^*/a^*)$$

with

$$a^{*2} = a_1 \cos \psi + b_1 \sin \psi$$

$$c^{*2} = c_1 \cos \psi + d_1 \sin \psi$$

3. Momocs: Outline analysis using R

Preliminaries

Momocs is S4-oriented ([Chambers 1998](#)) which has many advantages in terms of usage and

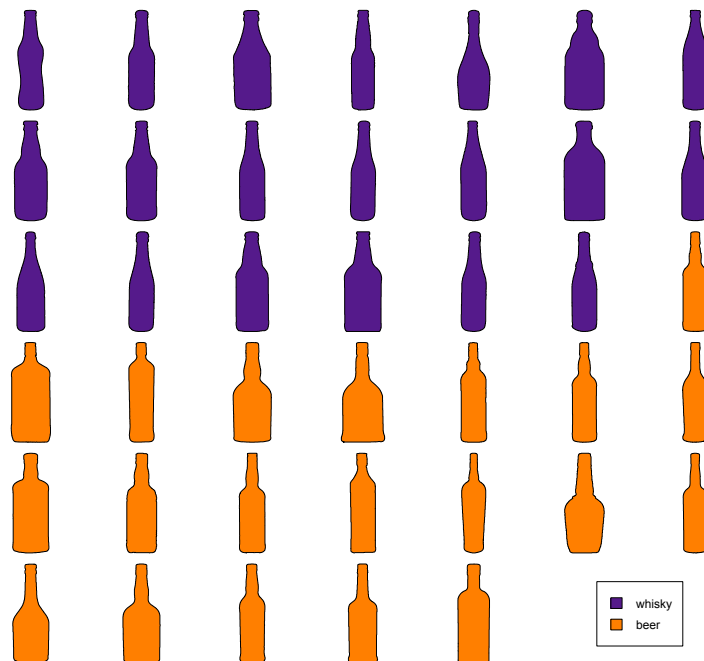


Figure 5: The bottles dataset included in **Momocs** and analyzed here.

programming: it prevents typing errors, provides validity checking, allows inheritance and encapsulation, etc. (see [Genolini 2008](#)). In practical terms, lists of coordinates and matrices of harmonic coefficients, are handled through ‘Coo’ and ‘Coe’ class objects respectively, to which methods can be applied. For those not familiar with S4 objects, data stored in objects can be retrieved and used as classical S3 objects in R: matrices, factors, etc. The **Momocs** documentation provides an extensive description of these classes and the methods that can be applied to them. The case study presented below will focus on the basic (and probably typical) use of the package.

All the following examples are based on the bottles dataset from the package (see `?bot` and Figure 5). We want to test if whisky and beer bottles have different shapes. How to calibrate outline analysis parameters and then obtain a matrix of normalized harmonic coefficients will also be discussed. While elliptic Fourier analysis will be presented because it is one of the most popular outline analysis tools today, the methodology employed will be equally valid for other approaches as well. On the extracted harmonic coefficients, some multivariate analyses will be presented: principal component analysis and morphological space, to illustrate the global bottle diversity, and multivariate ANOVA, to test for shape difference between the two sets of bottles. Then, linear discriminant analysis and hierarchical clustering will be introduced as perspectives for **Momocs** and because they are common and helpful statistical tools for those interested in multiple comparison. Finally, thin plate splines analysis will be introduced: this is not only a tribute to D’Arcy Thompson’s work but it may also bring great insights into the developmental differences underlying differences in the shapes compared.

3.1. Outline extraction

The input data for the morphometric analysis of outlines is, for two-dimension shapes, a set of $(x; y)$ pixel coordinates sampled on each outline. **Momocs** provides facilities to convert a set of .jpg black-and-white images, e.g., as classically obtained with image analysis stations when working on biological objects (see `?import.jpg`). Other image formats will be included in subsequent releases. The algorithm presented in Rohlf (1990), and implemented in R by Claude (2008) is used to extract the list of coordinates from a black mask that corresponds to the shape on a white background. Lists of coordinates, stored in separate .txt files, will be a convenient alternative when data are derived from GIS (see `?import.txt`).

Outlines are finally included in a 'Coo' class object. Outlines can be visualized in a one page graph (Figure 5). They can be centered, aligned, scaled and homologous landmarks can be defined to perform a Procrustes alignment (see Friess and Baylac 2003) before an elliptical Fourier analysis. When the outlines become rough due to artifacts during the digitization process (for instance when automatic outlining produces noise around the outline), outlines can be smoothed either when outlines are extracted from images, or before the calculation of harmonic coefficients (see `?coo.smooth` and `?eFourier` for instance). In order to specify explanatory variables going along with the coordinate or coefficient set, grouping factors or covariates can be specified through a `data.frame`, and then used to create subsets (see `?Coo`). Finally, we emphasize that **Momocs** supports import/export from/to other formats used in morphometrics such as .chc, .nef, etc.

```
R> library("Momocs")
R> data("bot", package = "Momocs")
R> panel(bot, cols = rep(col.sari(2), each = 20))
R> bot
```

A Coo object (see `?Coo`)

```
*****
```

General

```
-----
```

- 40 outlines
- 1864 +/- 371 coordinates per outline
- No landmark defined
- All outlines are closed
- 1 grouping factor(s) defined

Coordinates: @coo

```
-----
```

```
      [,1] [,2]
[1,]   37  561
[2,]   37  560
[3,]   37  559
[4,]   37  558

[...]
```

```
Factors: @fac
```

```
-----  
type : beer whisky
```

3.2. Calibration of outline analysis

Fourier-based approaches can fit any outline provided that the number of harmonics is large and the outline smooth enough between sampled points, while the signal/noise ratio can be very low for high order harmonics. The latter describe details that may be due to many things, i.e., digitalization artifacts or user bias, but not to real differences between shapes. On the other hand morphometrics is also used when differences between shapes are subtle. Conflicting situations such as those suggested by this familiar saying are actually a recurrent issue in morphometrics: what is the right number of harmonics? Unfortunately, no objective criterion exists so far, and the criterion used usually depends on the scope of the study. This might not be fully satisfactory to morphometrics newcomers but some approaches are presented below that can help to choose the most appropriate number of harmonics. Furthermore, a recent approach by [Claude \(2013\)](#) allows to study the measurement-error depending on the harmonic rank.

Through shape reconstruction

First, a ‘Coo’-object can be passed to `harm.qual()` to observe the reconstructed shape for a range of harmonics (Figure 6).

```
R> hqual(bot, method = "eFourier", id = 16, harm.range = 1:49,  
+       palette = col.sari, plot.method = "panel")
```

Through deviations

A qualitative approach would be of limited value, quite unlike the method we have presented to quantify deviations. The idea is to define, for a given number of sampled points, the best possible fit (i.e., obtained with half this number of points), and to then compare the euclidean distances obtained with a lower number of harmonics for every point of this outline and the best possible outline with these sampled points (Figure 7). One can for instance choose the minimal number of harmonics that leads to an average deviation of 1 pixel.

```
R> hquant(bot, harm.range = c(12, 16, 20, 24, 32))
```

Through harmonic power

Finally, we can also estimate the number of harmonics after examining the spectrum of harmonic Fourier power. The power is proportional to the harmonic amplitude and can be considered as a measure of shape information. As the rank of a harmonic increases, the power decreases and adds less and less information. We can evaluate the number of harmonics

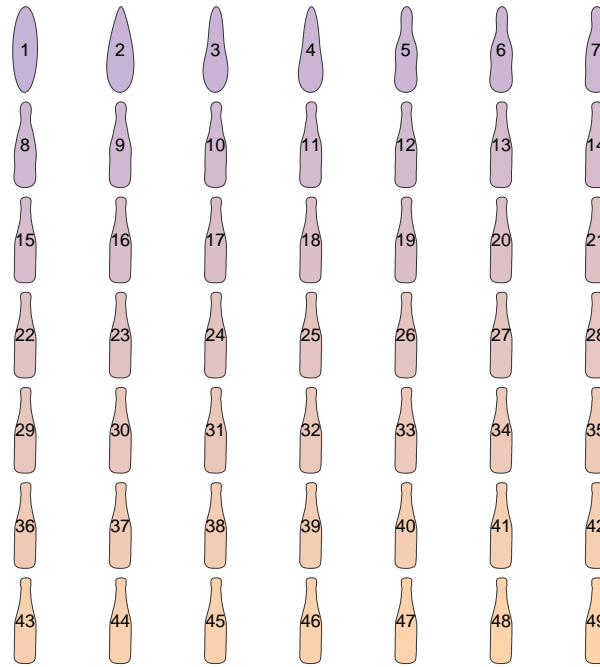


Figure 6: The “pecheresse” beer bottle in the **bot** dataset reconstructed from different numbers of harmonics. Twelve harmonics give a very satisfactory reconstruction and for 20 the result is almost perfect.

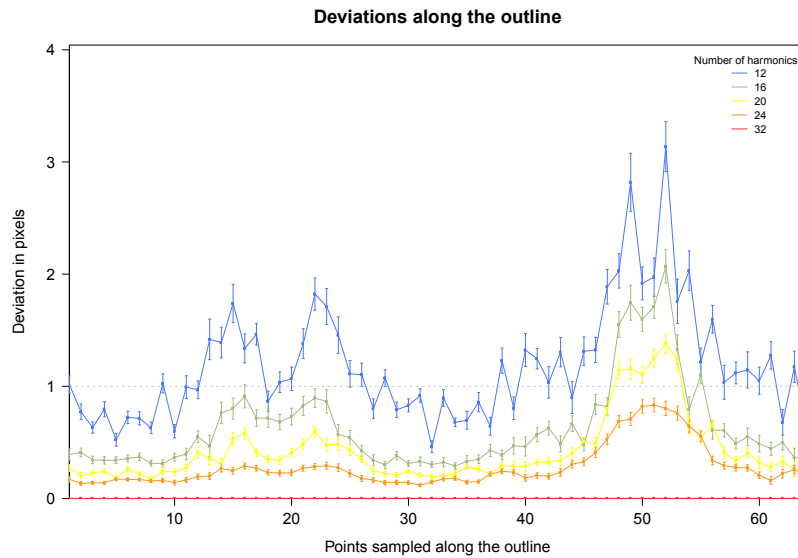


Figure 7: The bottles dataset included in **Momocs** and analyzed here. The y -axis represents the deviation in pixels between the best possible fit with a given number of harmonics (here 32) for every sampled points along the outline (on the x -axis). Standard deviations for the 40 shapes are displayed.

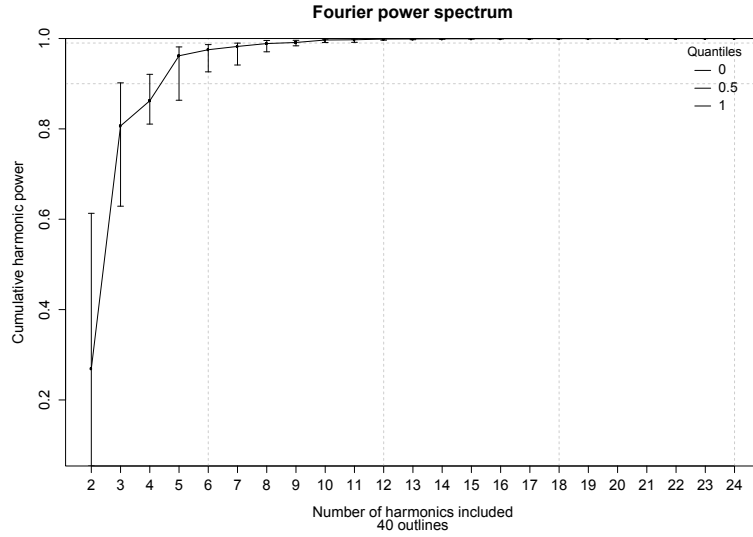


Figure 8: Cumulated harmonic Fourier power calculated from the bottles dataset. The 12 first harmonics gather nearly 100% of the harmonic power. Maxima and medians are also plotted.

that we must select, so their cumulative power gathers 99% of the total cumulative power (Crampton 1995, Figure 8). The power of a given harmonic is calculated as:

$$Power_n = \frac{A_n^2 + B_n^2 + C_n^2 + D_n^2}{2}$$

Figure 8 is obtained by typing:

```
R> hpow(bot)
```

Computing elliptic Fourier analysis

Once the right number of harmonics has been determined, elliptic Fourier analysis is performed on the ‘Coo’-object using the `eFourier` method and a

$$[\text{Number of Outlines}] \times [4 \text{ coefficients} \times \text{Number of Harmonics}]$$

matrix is obtained along with grouping factors, individual names, etc. and returned as a ‘Coe’ class object. By default, the obtained coefficients are normalized so that the first fitting ellipse is re-aligned along the x -axis. Other options can be considered, for instance, one can also normalize by performing a Procrustes alignment when landmarks can be defined (see Friess and Baylac 2003, `?ProcGPAAlign` and the package’s vignette which includes an illustration of this approach).

```
R> botF <- eFourier(bot, nb.h = 20)
R> botF
```

A matrix of harmonic coefficients obtained with
elliptical Fourier analysis (see ?Coe)

General

- 40 outlines
- 20 harmonics
- 1 grouping factor defined

[...]

Before multivariate analysis can be performed, one may be interested in having a global view of the elliptic Fourier analysis: which coefficients vary and what is the geometrical variation they depict (Figure 9). For instance, b_n and c_n represent the asymmetry of the shapes that can vary from coarse to fine scales, i.e., from lower to higher rank harmonics (Iwata, Niikura, Matsuura, Takano, and Ukai 1998; Yoshioka, Iwata, Ohsawa, and Ninomiya 2004). Figure 9 is obtained by typing:

```
R> hcontrib(botF, harm.range = 1:8)
R> boxplot(botF)
```

3.3. Analyzing Fourier coefficients

Principal component analysis (PCA) and other multivariate approaches can be directly performed on this ‘Coe’ class object (or directly on the matrix stored in the @coe slot) since all of the harmonic coefficients can be considered as quantitative variables. Both pros and cons of multivariate techniques apply to an analysis of the harmonic coefficients, just as they apply to other types of data. The purpose here is not describe them extensively but rather to present what is currently implemented in **Momocs**.

Principal component analysis

Momocs takes profit of the **ade4** package by Dray and Dufour (2007). The **pca** method can be used on a ‘Coe’ class object and performs a PCA with centering but no rescaling by default. In other words, the small-amplitude coefficients will contribute less than the first coefficients. It returns a ‘dudi’ object to which all suitable **ade4** functions can also be applied. An almost exhaustive wrapper that gathers graphical functions from **ade4** such as the display of eigenvalues, confidence ellipses and “stars”, individual labeling, neighboring graphs, etc. and that also adds dedicated features such as the display of the morphological space is provided by **dudi.plot**. See ?**dudi.pca** for an exhaustive description of this highly tunable function. Plotting the PCA (Figure 10) is straightforward: below, we first compute elliptical Fourier analysis with 20 harmonics, get out a ‘dudi’ object, and finally plot it.

```
R> botF <- eFourier(bot, nb.h = 20)
R> botD <- pca(botF)
```

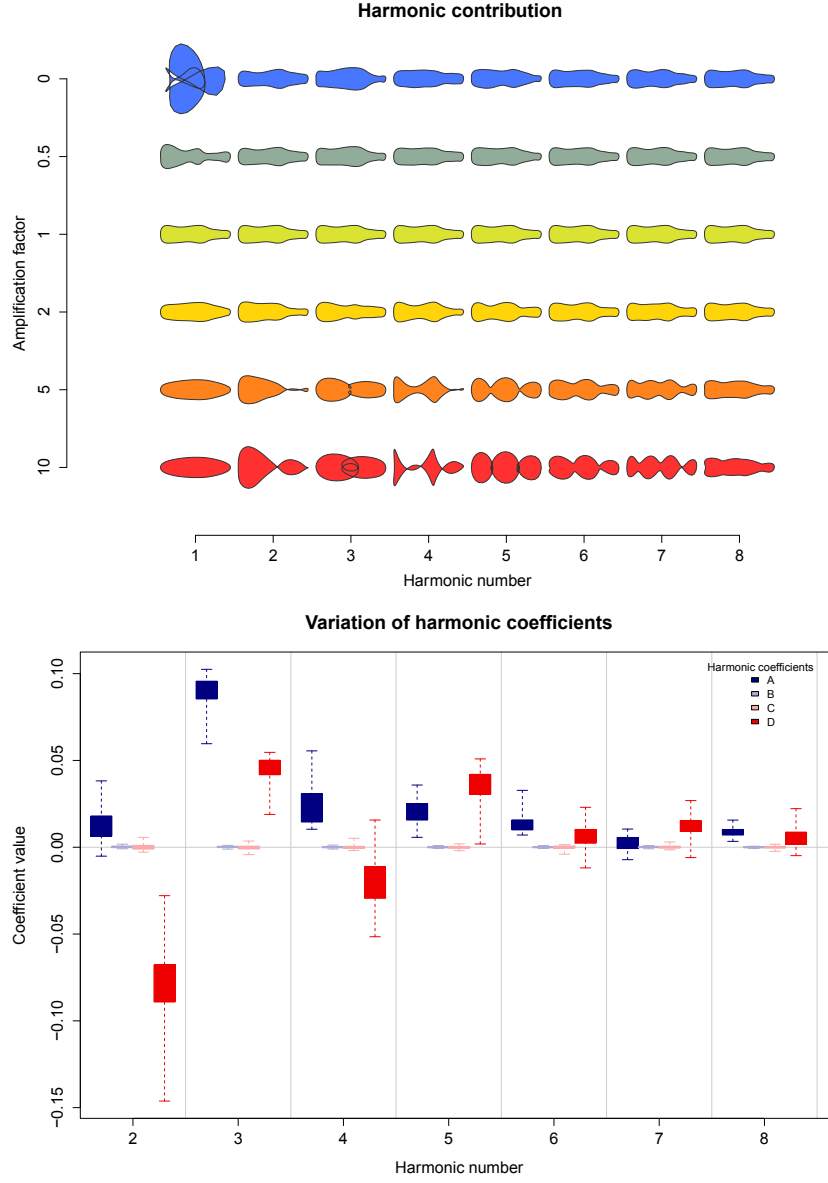



Figure 9: Analysis of harmonic coefficients after an elliptic Fourier computation on the `bot` dataset. The top figure shows the effect of every harmonic on the shape reconstruction. Every harmonic is represented on the x -axis; then the corresponding coefficients are multiplied by values that illustrate their removal (0), the normal shapes they lead to (1) or their exaggerated effect in shape reconstruction (2 and above). When the first harmonic is removed, on which all other ellipses “roll on”, the reconstructed shape is obviously very bad. Exaggerated coefficients help to understand their contribution to the final shape, *e.g.*, the second harmonic multiplied 5 or 10 times indicates that it contributes to describe the bottleneck while the third and fourth harmonic contribute to the constriction on the middle of the bottle. The bottom figure illustrates the variation of every coefficient along the whole dataset. Here, both the B_n and C_n coefficients are very small depicting the (bilateral) symmetry of the studied shapes.

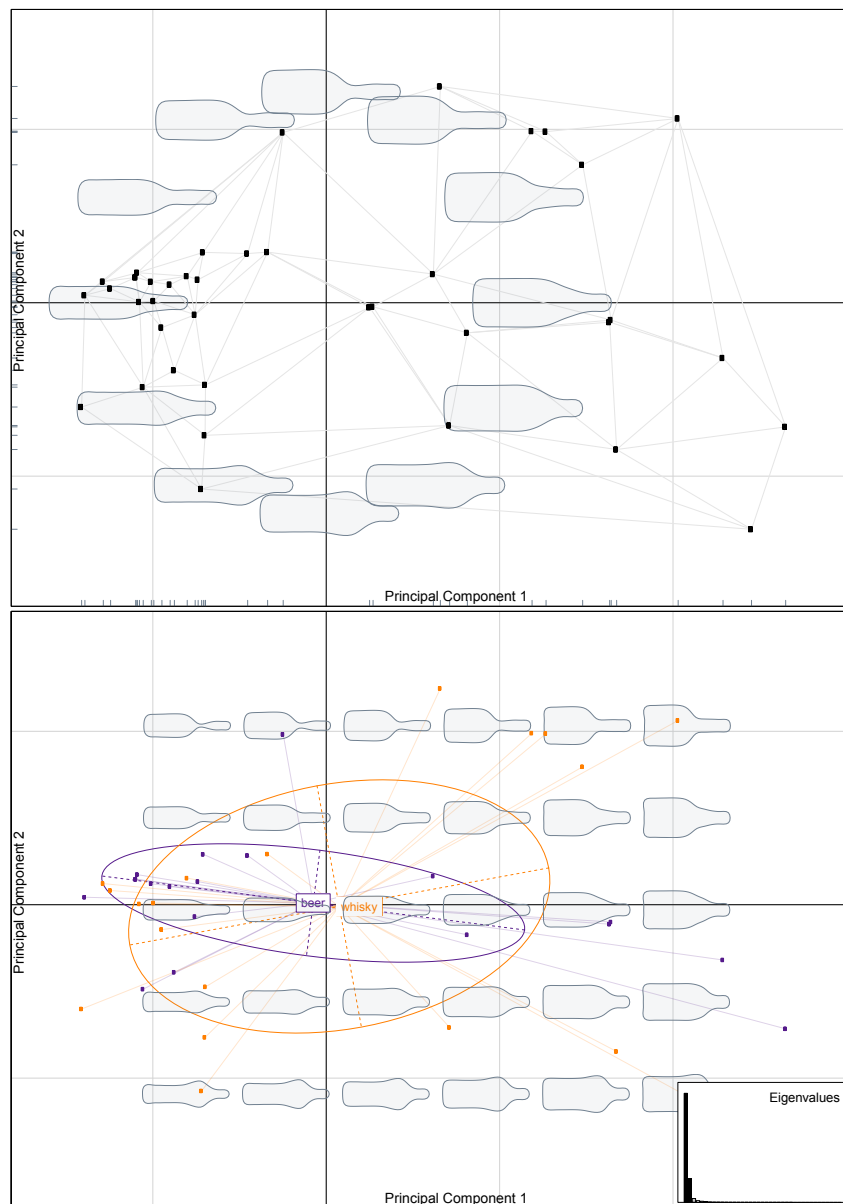


Figure 10: Some examples of the factorial maps depicting morphological variation that can be obtained with **Momocs**. The first two principal component axes are shown, (PC1 and PC2 are the x - and y -axis, respectively). The top figure displays the factorial map with no display of the classes, a neighboring graph, shapes reconstructed from the factorial map using the first two PC axes, and “rugs” along the axes; the bottom figure shows eigenvalues, confidence ellipses for the two groups and another option for displaying the morphological space, as a grid of bottles.

```
R> dudi.plot(botD, pos.shp = "circle", neighbors = TRUE)
R> dudi.plot(botD, 1, pch = 20, eigen = TRUE,
+   scale.shp = 0.5, amp.shp = 1.5, rug = FALSE)
```

Multivariate analysis of variance

We can test for a difference between subsets of shapes using multivariate analysis of variance (MANOVA), with every harmonic coefficient being considered as an homologous quantitative variable measured for every shape of the dataset. This can be achieved with:

```
R> manova.Coe(botF, "type")
```

The number or retained harmonics was not specified.

Analysis done with 9 harmonics

	Df	Hotelling-Lawley	approx F	num Df	den Df	Pr(>F)
fac	1	3826.1	318.84	36	3	0.000247 ***
Residuals	38					

Signif. codes: 0 "***" 0.001 "**" 0.01 "*" 0.05 "." 0.1 " " 1

The outcome of this analysis shows that shapes of the whisky and beer bottles significantly differ.

Hierarchical clustering

Momocs also includes a method to perform hierarchical clustering that hinges on `dist` and `hclust` for calculation, and `phylo.plot` from the **ape** package for graphical output (see [Paradis, Claude, and Strimmer 2004](#) and [Paradis 2012](#)). This can be achieved with the code below (Figure 11):

```
R> clust(botF)
```

Call:

```
hclust(d = dist(botE@coe, method = method))
```

```
Cluster method : complete
Distance       : euclidean
Number of objects: 40
```

Other methods, such as linear discriminant analysis, etc. exemplified in [Claude \(2013\)](#) will be integrated in further versions of **Momocs**.

Thin plate splines

Deformation grids as those that contributed to popularizing D'Arcy Thompson's ideas can be obtained using thin plate splines mathematical formalization. The notion of thin plate splines has been borrowed from mechanics and involves the bending of a thin sheet of metal (see [Bookstein 1991](#)). The deformations required to pass from the mean shape to the extreme points of the morphological space can be calculated and displayed on the PCA. One can also perform thin plate splines analysis based on the harmonic coefficients and the reconstructed shapes (Figure 12).

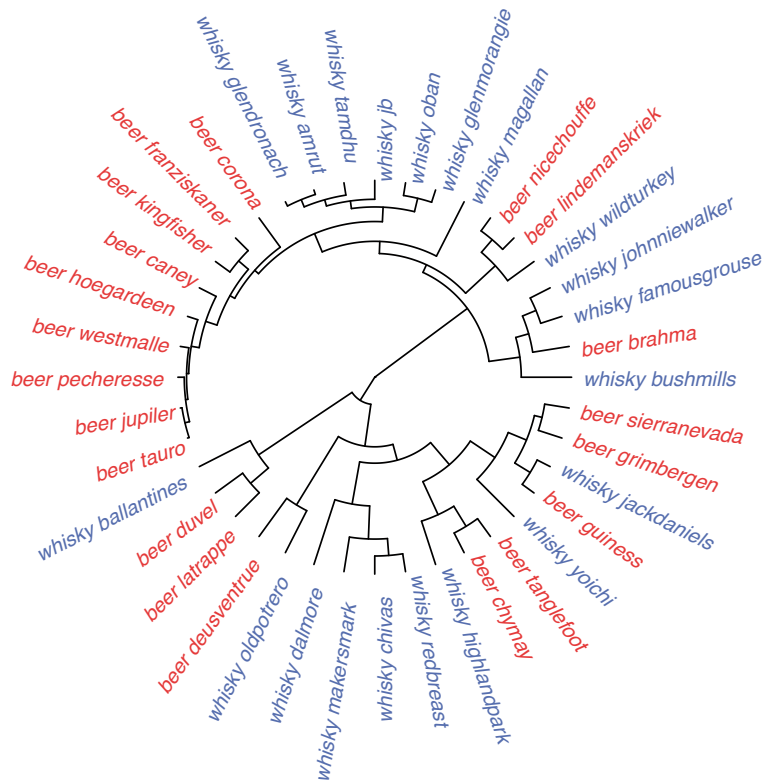


Figure 11: An example of the hierarchical clustering and the graphical output that can be obtained with **Momocs** on the matrix of coefficients calculated on the **bot** dataset.

```
R> botFg <- meanShapes(botF)
R> tps.grid(botFg$beer, botFg$whisky)
R> tps.arr(botFg$beer, botFg$whisky, amp = 2, arr.nb = 500,
+   palette = col.sari)
R> tps.iso(botFg$beer, botFg$whisky, iso.nb = 2000, amp = 2)
```

4. Perspectives for Momocs

The R package **Momocs** is generic enough to become a gateway in analysis of outline variation outside the developmental biology and evolutionary biology for which it has been developed. Complementary techniques can easily be included, by the package's developers upon request or by third parties who have developed new mathematical approaches. **Momocs** stands for *modern morphometrics* but, so far, only deals with outline analysis. For instance, we plan alongside package updates, a better digitalization step including outline acquisition with the help of Bezier curves, complementary approaches of outline sampling such as local over-sampling and additional smoothing algorithms, the integration of 3D algorithms and more gateways between **Momocs** and others morphometric programs.

Moreover, morphometric data, extracted as lists of coordinates are scarcely available on the web, but may be very useful for meta-analysis, development, and as support for course material. **Momocs** contains such datasets kindly provided by its users. We believe in such open

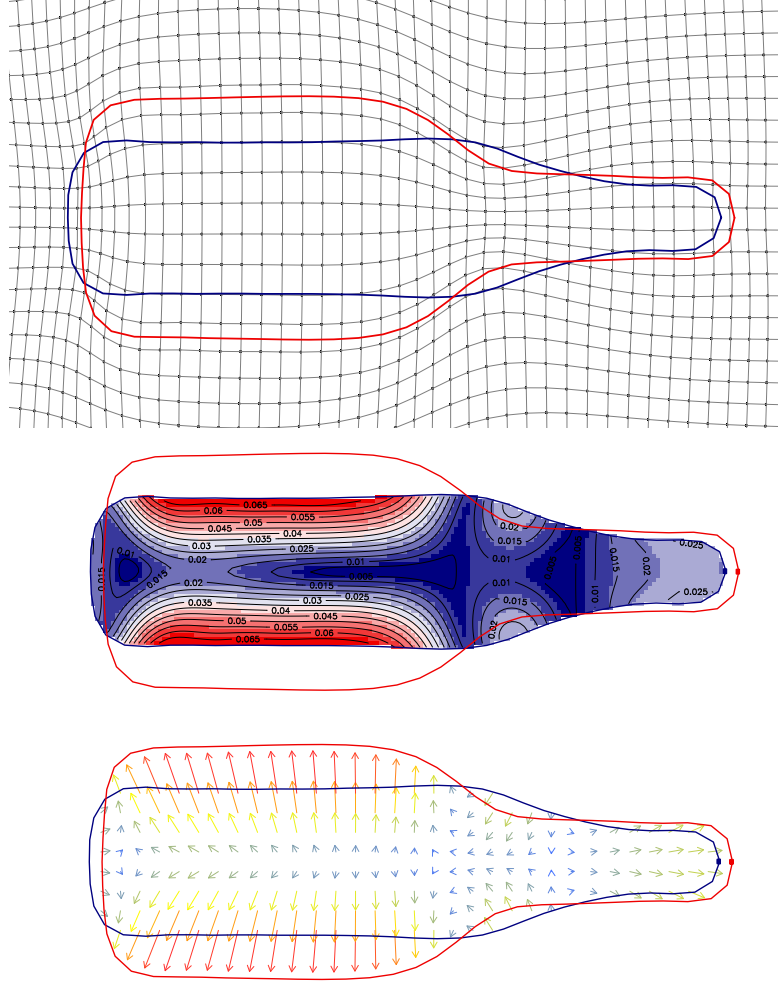


Figure 12: Some examples of the visualization of the deformation grids obtained using thin plate splines in **Momocs**. From top to bottom: a simple deformation grid; isodeformation lines; and a vector field that all depict the bindings required to pass from the average shapes for beer and whisky.

data philosophy, as encouraged by the R Core Team and the many package developers, and hope that **Momocs** will become a hub for such activity.

Acknowledgments

We thank the reviewers for all of their work and valuable comments made during the review process. We warmly thank Allan Bailur for his invaluable help with the presentation in English, along with Anna Seastrand on earlier versions of this manuscript. ANR-BDIV-07-008 supported VB and CG for this study; ANR-11-CEPL-002 and ANR-11-BCV-008 supported JC for this study.

References

- Adams D, Otárola-Castillo E (2012). *geomorph: Geometric Morphometric Analysis of 2d/3d Landmark Data*. R package version 1.0, URL <http://CRAN.R-project.org/package=geomorph>.
- Adams DC, Rohlf FJ, Slice DE (2004). “Geometric Morphometrics: Ten Years of Progress Following the ‘Revolution’.” *Italian Journal of Zoology*, **71**(1), 5–16.
- Bonhomme V (2012). “A Graphical Introduction to **Momocs** and Outline Analysis Using R.” Version 2012-10-17, URL <http://www.vincentbonhomme.fr/Momocs/vignettes/>.
- Bonhomme V, Frelat R, Gaucherel C (2013a). “Application of Elliptical Fourier Analysis to Watershed Boundaries: A Case Study in Haiti.” *Geomorphologie – Quantitative Hydro-Geomorphology*, **1**, 17–26.
- Bonhomme V, Prasad S, Gaucherel C (2013b). “Intraspecific Variability of Pollen Morphology as Revealed by Elliptic Fourier Analysis.” *Plant Systematics and Evolution*, **299**(5), 811–816.
- Bookstein FL (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press.
- Bowman A (2009). “Review of ‘Morphometrics with R’.” *Journal of Statistical Software, Book Reviews*, **31**(1), 1–2. Book author: Julien Claude, URL <http://www.jstatsoft.org/v31/b01/>.
- Chambers JM (1998). *Programming with Data: A Guide to the S Language*. Springer-Verlag.
- Claude J (2008). *Morphometrics with R*. Springer-Verlag, New York.
- Claude J (2013). “Log-Shape Ratios, Procrustes Superimposition, Elliptic Fourier Analysis: Three Worked Examples in R.” *Hystrix, the Italian Journal of Mammalogy*, **24**(1), 94–102.
- Courtiol A, Ferdy JB, Godelle B, Raymond M, Claude J (2010). “Height and Body Mass Influence on Human Body Outlines: a Quantitative Approach Using an Elliptic Fourier Analysis.” *American Journal of Physical Anthropology*, **142**(1), 22–29.
- Crampton JS (1995). “Elliptical Fourier Shape Analysis of Fossil Bivalves: Some Practical Considerations.” *Lethaia: An International Journal of Palaeontology and Stratigraphy*, **28**(2), 179–186.
- Dray S, Dufour AB (2007). “The **ade4** Package: Implementing the Duality Diagram for Ecologists.” *Journal of Statistical Software*, **22**(4), 1–20. URL <http://www.jstatsoft.org/v22/i04/>.
- Dryden I (2012). *shapes: Statistical Shape Analysis*. R package version 1.1-6, URL <http://CRAN.R-project.org/package=shapes>.
- Ferson S, Rohlf FJ, Koehn RK (1985). “Measuring Shape Variation of Two-Dimensional Outlines.” *Systematic Biology*, **34**(1), 59–68.

- Friess M, Baylac M (2003). “Exploring Artificial Cranial Deformation Using Elliptic Fourier Analysis of Procrustes Aligned Outlines.” *American Journal of Physical Anthropology*, **122**(1), 11–22.
- Genolini C (2008). *A (Not So) Short Introduction to S4*. URL <http://CRAN.R-project.org/doc/contrib/Genolini-S4tutorialV0-5en.pdf>.
- Giardina CR, Kuhl FP (1977). “Accuracy of Curve Approximation by Harmonically Related Vectors with Elliptical Loci.” *Computer Graphics and Image Processing*, **6**(3), 277–285.
- Iwata H, Nesumi H, Ninomiya S, Takano Y, Ukai Y (2002). “Diallel Analysis of Leaf Shape Variations of Citrus Varieties Based on Elliptic Fourier Descriptors.” *Breeding Science*, **52**(2), 89–94.
- Iwata H, Niikura S, Matsuura S, Takano Y, Ukai Y (1998). “Evaluation of Variation of Root Shape of Japanese Radish (*Raphanus Sativus* L.) Based on Image Analysis Using Elliptic Fourier Descriptors.” *Euphytica*, **102**(2), 143–149.
- Kendall D (1989). “A Survey of the Statistical Theory of Shape.” *Statistical Science*, **4**(2), 81–120.
- Klingenberg CP (2011). “**MorphoJ**: An Integrated Software Package for Geometric Morphometrics.” *Molecular Ecology Resources*, **11**(2), 353–357.
- Kuhl FP, Giardina CR (1982). “Elliptic Fourier Features of a Closed Contour.” *Computer Graphics and Image Processing*, **18**(3), 236–258.
- Macleod N (1999). “Generalizing and Extending the Eigenshape Method of Shape Space Visualization and Analysis.” *Paleobiology*, **25**(1), 107–138.
- Miller HJ, Wentz EA (2003). “Representation and Spatial Analysis in Geographic Information Systems.” *The Annals of the Association of American Geographers*, **93**(3), 574–594.
- Moellering H, Rayner JN (1981). “The Harmonic Analysis of Spatial Shapes Using Dual Axis Fourier Shape Analysis (DAFSA).” *Geographical Analysis*, **13**(1), 64–77.
- Paradis E (2012). *Analysis of Phylogenetics and Evolution with R*. 2nd edition. Springer-Verlag, New York.
- Paradis E, Claude J, Strimmer K (2004). “**ape**: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics*, **20**(2), 289–290.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Renaud S, Michaux JR (2003). “Adaptive Latitudinal Trends in the Mandible Shape of *Apodemus* Wood Mice.” *Journal of Biogeography*, **30**(10), 1617–1628.
- Richtsmeier JT, Burke Deleon V, Lele SR (2002). “The Promise of Geometric Morphometrics.” *American Journal of Physical Anthropology*, **119**(S35), 63–91.
- Richtsmeier JT, Cheverud JM, Lele S (1992). “Advances in Anthropological Morphometrics.” *Annual Review of Anthropology*, **21**, 283–305.

- Rohlf F, Archie J (1984). “A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae).” *Systematic Biology*, **33**(3), 302–317.
- Rohlf FJ (1990). “An Overview of Image Processing and Analysis Techniques for Morphometrics.” In *Proceedings of the Michigan Morphometrics Workshop*, Special Publication Number 2, pp. 47–60. University of Michigan Museum of Zoology, Ann Arbor.
- Rohlf FJ, Marcus LF (1993). “A Revolution in Morphometrics.” *Trends in Ecology and Evolution*, **8**(4), 129–32.
- Rohlf FJ, Slice D (1990). “Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks.” *Systematic Zoology*, **39**(1), 40–59.
- Small CG (1996). *The Statistical Theory of Shape*. Springer-Verlag.
- Thompson DW (1917). *On Growth and Form*. Dover.
- Wentz EA (2000). “A Shape Definition for Geographic Applications Based on Edge, Elongation, and Perforation.” *Geographical Analysis*, **32**(2), 95–112.
- Yoshioka Y, Iwata H, Ohsawa R, Ninomiya S (2004). “Analysis of Petal Shape Variation of *Primula sieboldii* by Elliptic Fourier Descriptors and Principal Component Analysis.” *The Annals of Botany*, **94**(5), 657–664.
- Zahn CT, Roskies RZ (1972). “Fourier Descriptors for Plane Closed Curves.” *IEEE Transactions on Computers*, **21**(3), 269–281.

Affiliation:

Vincent Bonhomme
Department of Ecology
French Institute of Pondicherry
600 001 Pondicherry, India
E-mail: Vincent.Bonhomme@ifpindia.org

Current affiliation:

Centre de Bio-Archéologie et d'Ecologie
UMR 5059 CBAE
34 000 Montpellier, France

Sandrine Picq
Centre de Bio-Archéologie et d'Ecologie
UMR 5059 CBAE
34 000 Montpellier, France
Current affiliation:
Université du Québec en Abitibi-Témiscamingue
Rouyn-Noranda J9X 5E4, Québec, Canada

Julien Claude
Institut des Sciences de l'Évolution
Université de Montpellier II
34 095 Montpellier, France

Cédric Gaucherel
Department of Ecology
French Institute of Pondicherry
600 001 Pondicherry, India