Functional data analysis in shape analysis

Irene Epifanio [a,*], Noelia Ventura-Campos [b]

[a] *Dept. Matemàtiques, Universitat Jaume I, Campus del Riu Sec, 12071 Castelló, Spain*
[b] *Dept. Psicologia Bàsica, Clínica i Psicobiologia, Universitat Jaume I, Spain*

ARTICLE INFO

ABSTRACT

Mid-level processes on images often return outputs in functional form. In this context the use of functional data analysis (FDA) in image analysis is considered. In particular, attention is focussed on shape analysis, where the use of FDA in the functional approach (contour functions) shows its superiority over other approaches, such as the landmark based approach or the set theory approach, on two different problems (principal component analysis and discriminant analysis) in a well-known database of bone outlines. Furthermore, a problem that has hardly ever been considered in the literature is dealt with: multivariate functional discrimination. A discriminant function based on independent component analysis for indicating where the differences between groups are and what their level of discrimination is, is proposed. The classification results obtained with the methodology are very promising. Finally, an analysis of hippocampal differences in Alzheimer's disease is carried out.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Functional data analysis (FDA) provides statistical procedures for functional observations (a whole function is a datum). The goals of functional data analysis are basically the same as those of any other branch of statistics. Ramsay and Silverman (2005) give an excellent overview. Ferraty and Vieu (2006) provide a complementary and very interesting view on nonparametric methods for functional data. The field of FDA is quite new and there is still a lot of work to be done, but in recent years several applications have been developed in different fields (Ramsay and Silverman, 2002). Furthermore, the software that the authors used is available on the website for those books. The wide variety of disciplines where FDA is applied is also shown through the Special Issue on Statistics for Functional Data (González-Manteiga and Vieu, 2007) published in this journal in 2007. Some examples of these fields of application are: climatology, chemicals, geophysics and oceanology, economics, remote sensing, demographics (Delicado, 2011), materials science (Berrendero et al., 2011), biostatistics or genetics (López-Pintado and Romo, 2011; Li and Chiou, 2011). A mixture of practical and theoretical aspects is found in Ferraty and Romain (2011).

Many mid-level processes on images return outputs in functional form, such as granulometries and other morphological curves (Soille, 2003), or spectral-energy descriptors (González et al., 2004, p. 468). Although FDA techniques are specifically designed to deal with functions and are natural tools for their analysis, FDA has hardly been used in image analysis (to the best of our knowledge, the only article that uses FDA for analyzing 2 dimensional profiles is Nettel-Aguirre, 2008). The reason for this could be because FDA is quite new (the first book on FDA was published in 1997 (Ramsay and Silverman, 2005)), and it is not very well known in the image processing community.

---

* Corresponding author. Tel.: +34 964728390; fax: +34 964728429.
  E-mail address: epifanio@uji.es (I. Epifanio).

Shape analysis is a field where functions are frequently used to represent shape (Kindratenko, 2003). According to Stoyan and Stoyan (1994), shapes can be described by three kinds of tools: first, set descriptors and mathematical morphology tools; second, using landmarks (point description); and third, employing a function describing the contours (see Kindratenko (2003) for a review of various contour functions and methods for their analysis).

One of the objectives of this study is to highlight the advantages of the use of FDA in image analysis, and particularly in shape analysis. In Section 2 we analyze a well-known database of bones with these three approaches (the set theory approach, the landmark based approach and the functional approach) and compare their results in two of the main problems in form statistics (Stoyan and Stoyan, 1994): the study of the main sources of variation among the shapes (principal component analysis), and classification among different classes (discriminant analysis). The analysis of contour functions by FDA gives more meaningful results. The contour functions used in Section 2 are multivariate (two functions define the contours). Recently, independent component analysis (ICA) has been successfully used for the classification of univariate curves (Epifanio, 2008). This methodology is extended to the multivariate case in Section 3.2, where a discriminant function based on ICA is also introduced. Discriminant results obtained with this methodology are very promising (in Section 3.2 a small comparative review is also given). Section 4 shows a FDA application in the analysis of magnetic resonance (MR) scans in order to study the hippocampal differences among the subjects of three groups: controls, patients with mild cognitive impairment (MCI), and patients with early AD (Alzheimer's disease). Finally, conclusions and some open problems are discussed in Section 5.

2. Shapes through three approaches

In order to compare the three perspectives, we decide to use a well-known and extensively studied database, which is publicly available: the bone shapes from a paleopathology study analyzed in Ramsay and Silverman (2002, Chapter 4). Furthermore, this database has been analyzed previously in the literature using both landmarks and images, so it is perfect for our purposes. A total of 68 outlines were studied, which correspond to 52 non-eburnated and 16 eburnated femora. They considered 12 landmarks (see Shepstone et al., 1999, for details about data and identification and construction of these landmarks).[1]

Apart from the landmarks, the images from which they were extracted are also available[2] (see Shepstone et al., 2000, for details about the data). As explained in the file, these images are not binary, but they can be easily binarized because the background (zone without bone) is coded as zero. Although there are 121 images in that file, we only consider the images corresponding to the 68 cases available for the landmarks. Each image is rotated so that the 'bottom' of the condyles sits on the bottom of the image (as explained in Shepstone et al. (2000), each femur was rotated in the horizontal plane until the articular surface was parallel with the plane of the camera lens). Furthermore, all left femora were reflected to produce 'right' images so that the left side of any image indicates the lateral side and the right side the medial.

2.1. Landmark description

Each individual outline yields a vector of 24 coordinates (12 landmarks in 2 dimensions). In order to analyze the sample of 68 outlines, we follow the approach explained in Dryden and Mardia (1998, Chapter 5). The library *shapes* (Dryden, 2007) of free software R (R Development Core Team, 2010) provides the routines for this analysis.

First, a generalized Procrustes analysis is carried out to register landmark configurations into optimal registration using translation, rotation and scaling. Note that in Ramsay and Silverman (2002, Chapter 4) an ordinary Procrustes analysis was carried out, where two configurations were matched (each configuration is matched to the original mean configuration). However, we use a generalized Procrustes analysis (*procGPA* with default arguments), which was developed to match configurations when more than two objects are available, as in our case. Therefore, our results do not coincide with those presented in Ramsay and Silverman (2002, Chapter 4). Fig. 1 shows the Procrustes mean shape scaled to unit size. We standardize for size by dividing by the centroid size, which is our measure of size (see Dryden and Mardia (1998, Chapter 2) for details). The centroid size is the square root of the sum of squared distances from each landmark to the centroid.

Second, the routine *procGPA* also returns the principal component analysis. To visualize the effect of each principal component (PC) easily, we draw vectors from the mean shape to a shape at +6 standard deviations along the PCs, except for the first PC where we use +2 standard deviation. In Fig. 2 we can see these vectors.

The first four principal components explain 92.76% in total: 84.62%, 3.26%, 2.96% and 1.92% of variability, respectively. There appears to be a high dependence between certain landmarks, as indicated by the fact that the first PC explains such a large proportion of the variability. The first PC includes the movement of many landmarks outwards, whereas there is less movement in the intercondylar notch (landmarks 6, 7 and 8), and especially in landmark 7. On the other hand, the second PC involves a shift inwards for landmarks 6 and 7. The third and fourth PCs show more complex, asymmetrical movement in landmarks.

---

[1] Available together with the code on website: http://www.stats.ox.ac.uk/~silverma/fdacasebook/boneshapechap.html.

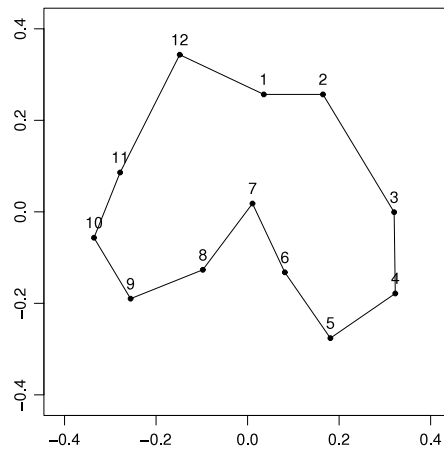[2] Available on website: http://www.stats.ox.ac.uk/~silverma/data/bones.tar.gz.

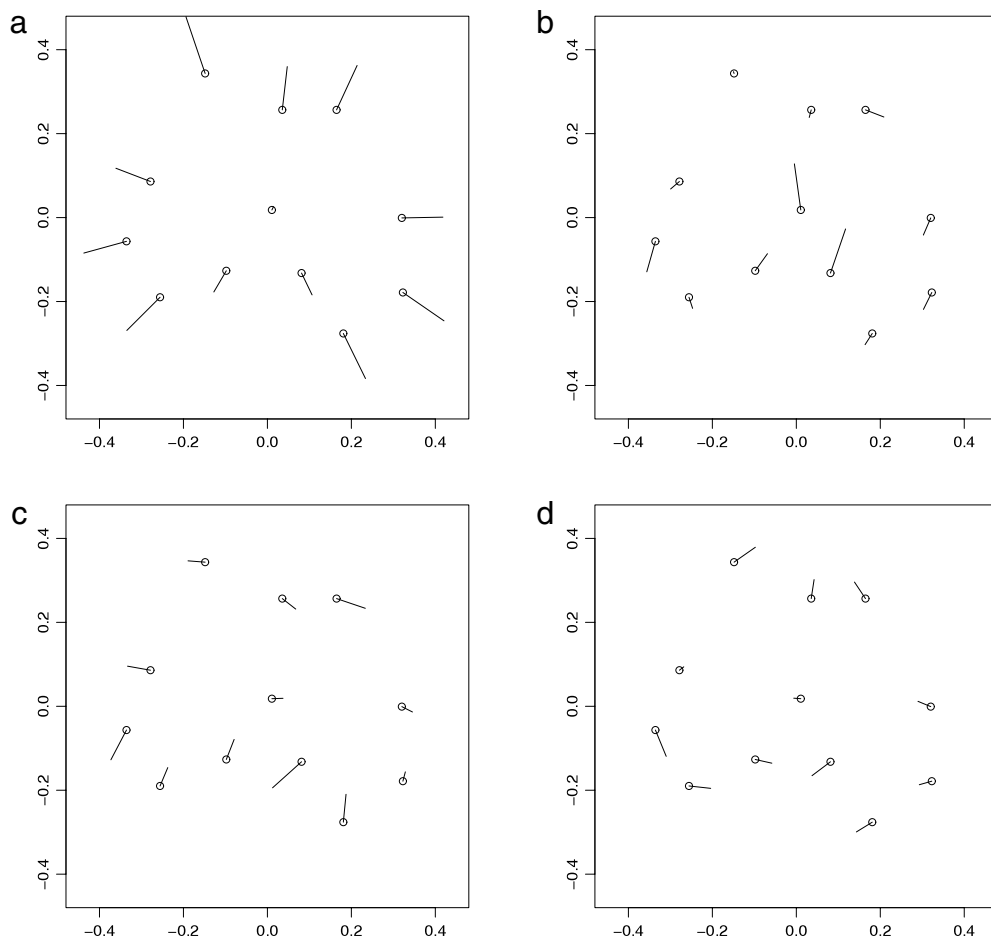**Fig. 1.** Procrustes mean shape of bones, with 12 landmarks used.

Fig. 2. The mean shape with vectors (see text for details) along the first PC (a), the second PC (b), the third PC (c) and the fourth PC (d).

We are also interested in discovering how arthritic bones differ from controls. In Ramsay and Silverman (2002, Chapter 4), for the principal component scores, a t-test was carried out to compare the eburnated and non-eburnated bones. As explained in Jolliffe (2002, Chapter 9), PCA can be used in discriminant analysis in order to reduce the dimensionality of the analysis, assuming the covariance matrix is the same for all groups. However, we have to be aware that there is no guarantee that the separation between groups will be in the direction of the high-variance PCs, the separation between groups may be in the directions of the last few PCs. On the other hand, each PCs contribution in linear discriminant analysis can be assessed
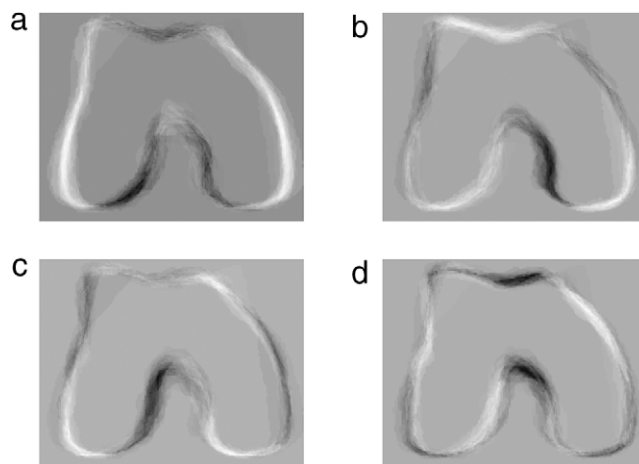
Fig. 3. Loadings in the first PC (a), the second PC (b), the third PC (c) and the fourth PC (d), for the set approach.

independently thanks to their uncorrelatedness. We compute all PCs, and we consider only those for which the difference (between eburnated and non-eburnated bones) is significant with $\alpha = 0.05$, that is to say, when the p-value of the t-test is less than 0.05.

The difference is only significant on component 3 of 24 (p-value $= 0.002$). Using leave-one-out cross validation with linear discriminant analysis (lda function from library MASS Venables and Ripley, 2002) for the scores of this component, 16 errors are obtained: 2 false positives (non-eburnated bones that are classified as eburnated) and 14 false negatives (eburnated bones that fail to be so classified). If we use component 3 together with component 7, which is the following component with the smallest p-value (0.052), the number of errors is the same (16 errors), with 4 false positive and 12 false negative. Note that this is an underestimation of the true error rate, since we are computing PCA and choosing the components with all samples; even so, this error rate will be worse than those that we will obtain with the other perspectives. This was predictable because more information (not just 12 points) about the shapes will be used in other perspectives.

2.2. Set description

Before the analysis of the images, position and size information will be filtered out. Remember that 'left' images have already been reflected, and all them have already been rotated (see Shepstone et al., 2000). We remove the location by translating each image to the origin in such a way that its centroid coincides with the origin. Images are also standardized for size, in the same way as we did with the landmark approach. Scale is removed by dividing through the centroid's size. If $\{X_j\}$ ($j = 1, \ldots, k$) is the set of all the points in each digitalized figure, each point is divided by the centroid's size ($\sqrt{\sum_{j=1}^{k} \|X_j - \bar{X}\|^2}$, where $\bar{X}$ is the average or centroid, and $\|\cdot\|$ stands for the Euclidean norm).

As shapes are considered as sets, we compute the PCA as explained in Horgan (2000). Here, a variable X, which can only take value 1 or 0 depending on whether it belongs to the shape or not, is associated with each position of the image. We therefore do PCA for binary data. Note that PCA is equivalent to a Principal coordinate analysis (classical multidimensional scaling) of the matrix of Euclidean distances between the observations (see Horgan, 2000, for details). For that reason, according to Gower (1966), PCA can provide a plausible low-dimensional representation when all variables are binary.

The percentages of variance accounted for the first four principal components are 13.62%, 9.47%, 8.37% and 6.22% with a cumulative total of 37.67%. 44 components are necessary in order to capture 90.39% of the variability. The loadings in the first four principal components are represented as gray levels, as done in Horgan (2000), and appear in Fig. 3. Bright gray levels indicate positive loadings, and dark negative ones. An interpretation of them is not easy. As a tentative interpretation: the first component is greatly concentrated on the external part of the condyles, the second component on the right condyle, while the third component is concentrated on left condyle. On the other hand, the fourth component is associated with the joint between condyles and the top part of the image.

For the discriminant analysis, we follow the same strategy as in the landmark approach. We compute all PCs with all data, and we consider only those for which the difference (between eburnated and non-eburnated bones) is significant. Now, five components are significant: 2, 3, 41, 43 and 59. If we compute the misclassification rate by leave-one-out as before, the best classification with one component is obtained by the second component (15 errors, 3 false positive and 12 false negative). If two components are used, the best classification is achieved by components 2 and 3, giving 12 errors (4 false positive and 8 false negative). Although in this approach (and the following) more than two components could be considered jointly, we restrict ourselves to classifying with only two components, for the following reasons: with only two components we try to avoid an overfitted model, and in fact we obtain a simple and parsimonious model which is able to classify with

very good results; two components (dimensions) are easier to interpret and represent graphically; considering only pairs is computationally faster than considering all the possible subsets and finally, as only one component is significant in the landmark approach, the use of many more components in the other approaches would not be a fair comparison.

However, the error rates obtained before (15 and 12 for one and two components, respectively) are too optimistic, as we are using all the data for obtaining the PCs. As explained in Ripley (1996, Chapter 2) or Hastie et al. (2009, Chapter 7), a double or nested cross-validation should be done. The double cross-validated error rate estimation that we propose makes use of the cross-validation methodology in two stages in order to estimate the overall error rate of the whole procedure, also taking into account the feature-selection method, which must be implemented in the training phase (see Ambroise and McLachlan, 2002, for a detailed discussion of this point). First, each observation (bone) is deleted in turn (external cross-validation). For the remaining samples, we carry out the PCA and select those components for which the difference is significant. As mentioned previously, we restrict ourselves to classifying with only two components, therefore we selected the pair from the selected components that gives the smallest error rate by leave-one-out cross-validation (internal cross-validation) with linear discrimination for the remaining samples. Using the scores of the two selected components, we calculate the linear discriminant classifier with the internal samples. The scores of the initially deleted observation for these two chosen components can be easily obtained as we have their loadings; it is a simple inner product. We use the linear classifier previously built to predict the class label for the initially deleted observation. This prediction is preserved to produce the (external) leave-one-out estimate of the prediction error, because this process is repeated in turn for each of the observations. Of course, for each observation (external iteration), different subsets of PC can be selected. As PCA is computed with different data, PCs from different external iterations are not comparable, especially those with low variance. For example, component 20 in one iteration could be quite different from component 20 obtained in another iteration.

By the double cross-validation strategy, a total of 15 errors are obtained (3 false positive and 12 false negative).

2.3. Function description

With the standardized images (position and size filtered out) as explained in Section 2.2, we consider the contour (outline) of the shapes. The contour is a closed planar curve that consists of the elements of the figure boundary. Although other contour functions can be used (see Kindratenko, 2003, for a systematic review of various contour functions), we consider the contour parameterization by its arc length, which can be applied to any contour (note that other contour functions have limitations). The tracing begins counterclockwise at the easternmost outline point in the same row as the centroid, using *bwtraceboundary* from the image toolbox of MatLab. We normalize these functions in such a way that the perimeter length is eliminated and the functions are defined on [0,1]. Although they are recorded discretely, a continuous curve or function lies behind these data. In order to convert the discrete curve observations into a true functional form, we approximate (smooth) each curve by a weighted sum (a linear combination) of fifty-one Fourier bases (note that this basis system is periodic with period 1), and determine the coefficients of the expansion by fitting data by least squares, as explained in Ramsay and Silverman (2005, Chapter 4). Each curve is, therefore, completely determined by the coefficients in this basis, and each function is computable for any desired argument value $t \in [0, 1]$. All this work has been done by means of *fda* library (Ramsay and Silverman, 2005). The free library *fda* for MatLab and R, available at http://www.functionaldata.org, is especially designed to work with functional data (Ramsay et al. (2009) is a book about this library). Finally, in order to have the same number of points for all functions, we evaluate the functions in 100 equidistant points from 0 to 1. We therefore have two pairs of functions (representing coordinates) $\{X(t), Y(t)\}$ for each bone, with $t \in [0, 1]$.

Let us see how to apply PCA in this infinite dimensional domain. A short answer would be that summations change into integrations, but details are given in the following section.

2.3.1. PCA for functional data

In order to see how PCA works in the functional context, let us recall PCA for Multivariate Data analysis (MDA). In MDA, principal components are obtained by solving the eigenequation

$$\mathbf{V}\xi = \rho\xi, \tag{1}$$

where $\mathbf{V}$ is the sample variance–covariance matrix, $\mathbf{V} = (N-1)^{-1}\mathbf{X}'\mathbf{X}$, where in turn $\mathbf{X}$ is the centered data matrix, N is the number of individuals observed, and $\mathbf{X}'$ indicates the transpose of $\mathbf{X}$. Furthermore, ξ is an eigenvector of $\mathbf{V}$ and ρ is an eigenvalue of $\mathbf{V}$.

In the functional version of PCA, vectors are not considered any more, but PCs are replaced by functions or curves. Let $\{x_1(t), \ldots, x_N(t)\}$ be the set of observed functions. The mean function can be defined as the average of the functions point-wise across replications ($\bar{x}(t) = N^{-1} \sum_{i=1}^{N} x_i(t)$). Let us assume that we work with centered data (the mean function has been subtracted), and define the covariance function $v(s,t)$ analogously by $v(s, t) = (N-1)^{-1} \sum_{i=1}^{N} x_i(s)x_i(t)$. As explained in Ramsay and Silverman (2005, Chapter 8), the functional counterpart of Eq. (1) is the following functional eigenequation

$$\int v(s, t)\xi(t)\mathrm{d}t = \rho\xi(s), \tag{2}$$

where ρ is still an eigenvalue, but where $\xi(s)$ is an eigenfunction of the variance–covariance function, rather than an eigenvector. Now, the principal component score corresponding to $\xi(s)$ is computed by using the inner product for functions

$$s_i = \int x_i(s)\xi(s)\mathrm{d}s. \tag{3}$$

Note that for multivariate data, the index s is not continuous, but a discrete index j replaces it: $s_i = \sum_j x_{ij}\xi_j$.

There are several strategies for solving the eigenanalysis problem in Eq. (2). In order to retain the continuity of the original functional data and to reduce the amount of information, we have used the approach proposed in Ramsay and Silverman (2005). Instead of using a lot of variables obtained by discretizing the original functions, this type of analysis works with the coefficients of the functions expressed as a linear combination of known basis functions (Fourier in our case, although other bases could be used, such as B-splines). Functional PCA can be carried out easily by using the library *fda*. For a complete review of computational methods for functional PCA, see Ramsay and Silverman (2005).

Regarding the problem of how many PCs can be computed, let us note that in the functional context, "variables" now correspond to values of t, and there is no limit to these. Therefore, a maximum of $N - 1$ components can be computed. However, if the number of basis functions K defining the curves is less than N, K would be the maximum.

2.3.2. Functional PCA with multiple functions

Our data consist of two functional data per bone. Functional PCA can deal with two (or more) functional observations per individual, two curves $x(t)$ and $y(t)$. Let $\{(x_1(t), y_1(t)), \ldots, (x_N(t), y_N(t))\}$ be the set of pairs of observed functions. Two mean functions $(\bar{x}(t), \bar{y}(t))$ and two covariance functions $(v_{XX}(s, t), v_{YY}(s, t))$ can be computed for each kind of function, respectively. Furthermore, we can calculate the cross-covariance function of the centered data by:

$$v_{XY}(s, t) = (N - 1)^{-1} \sum_{i=1}^{N} x_i(s)y_i(t). \tag{4}$$

A typical PC is defined by a two-vector $\xi = (\xi_X, \xi_Y)$ of weight functions (two curves). They are solutions of the eigenequation system $V\xi = \rho\xi$, which in this case can be written as

$$\int v_{XX}(s, t)\xi_X(t)\mathrm{d}t + \int v_{XY}(s, t)\xi_Y(t)\mathrm{d}t = \rho\xi_X(s)$$
and
$$\int v_{XY}(s, t)\xi_X(t)\mathrm{d}t + \int v_{YY}(s, t)\xi_Y(t)\mathrm{d}t = \rho\xi_Y(s). \tag{5}$$

Now, the PC score for the i-th bivariate function $(x_i(t), y_i(t))$ is computed by $s_i = \int x_i\xi_X + \int y_i\xi_Y$, because the inner product between bivariate functions is defined by the addition of the inner products of the two components. This amounts to stringing two functions together to form a composite function.

To solve the eigenequation system, each function $x_i(t)$ and $y_i(t)$ is replaced by a vector of values or basis coefficients, and a single synthetic function is built by joining them together. When PCs have been computed, we separate the parts belonging to each coordinate. Again, this procedure is implemented on the *fda* library and is explained fully in Ramsay and Silverman (2005).

The proportion of variance explained by each eigenfunction is computed as in the multivariate case, by each eigenvalue ρ divided by the sum of all eigenvalues. Moreover, for each PC, the variation accounted for each original curve $x(t)$ and $y(t)$ is given by $\int \xi_X(s)\xi_X(s)\mathrm{d}s$ and $\int \xi_Y(s)\xi_Y(s)\mathrm{d}s$ respectively, because their sum is one by definition.

The first four principal components for the bones explain 80.52% of the whole variance, made up of 55.00%, 10.59%, 9.02% and 5.91% respectively. 90.21% of the variability is explained by the first seven components.

In order to display the effect of each PC, the mean function is displayed together with the functions obtained by adding (plotted with +) and subtracting (plotted with −) a suitable multiple of the principal component function in question, in our case two standard deviations of each component. In this way, the effects of each PC are usually clarified (see Ramsay and Silverman, 2005, Chapter 8). They are shown in Fig. 4. The first component corresponds to the length of the left condyle, but also to the shape of the intercondylar notch. 63.33% of the variation in this component is due to the y-coordinates (function). The second component is associated with a broader intercondylar notch (for negative scores) and the shape of the top of the image. In this component, 95.13% of variability comes from the x-coordinates. Component 3 is associated with the left condyle and the top right hand part of the image. The proportion of the variability in this component is 74.59% for the x-coordinates. Finally, component 4 is concentrated almost entirely on the internal part of the right condyle. 58.12% of the variation in this component is due to the y-coordinates.

For the discriminant analysis, we follow the same strategies as in the set approach. First, PCA with all data is calculated, but only those components for which the difference is significant are considered. Now, three components are significant: 1, 6 and 10. If we compute the misclassification rate by leave-one-out as before, the best classification with one component is obtained by the sixth component (16 errors, 3 false positive and 13 false negative) and the tenth component (16 errors, 0 false positive and 16 false negative). If two components are used, the best classification is achieved by components 1 and 6,
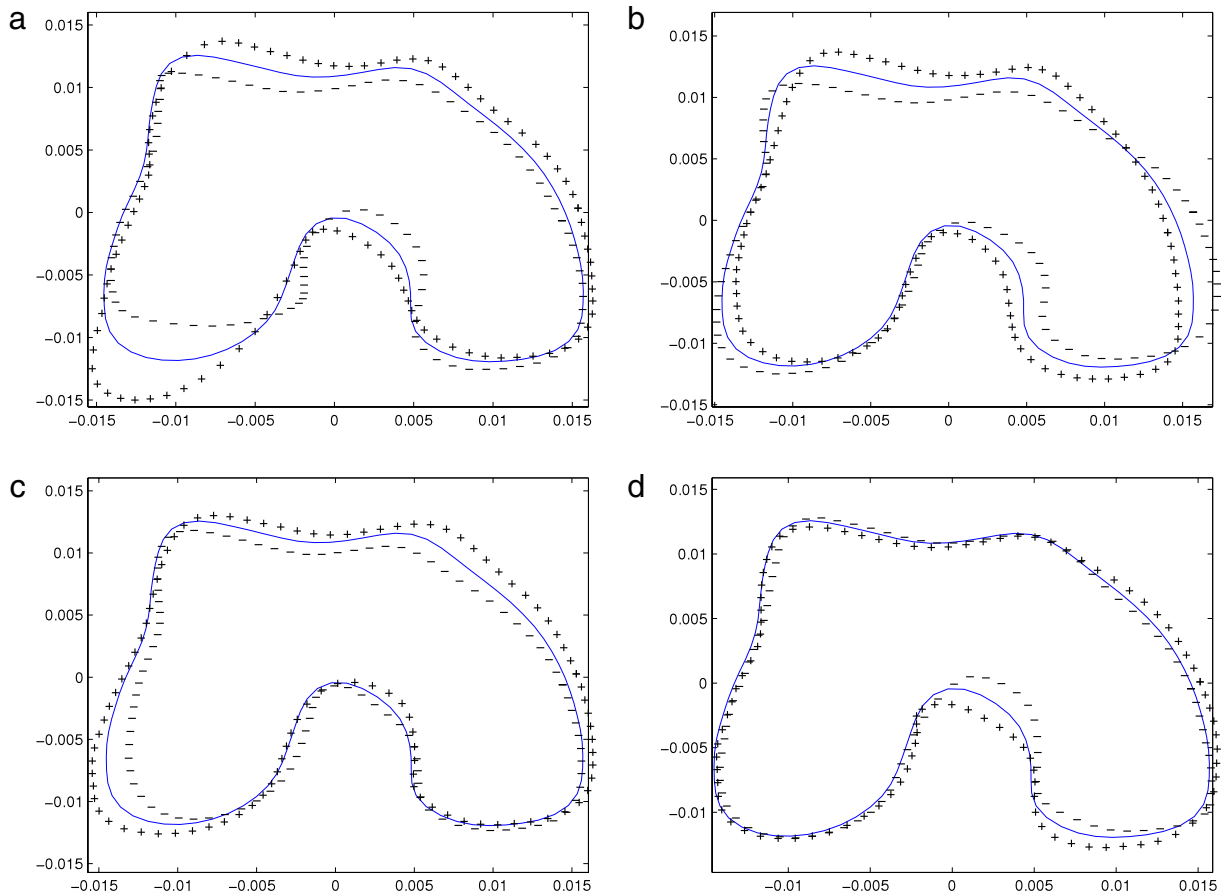
Fig. 4. The effects of the first PC (a), the second PC (b), the third PC (c) and the fourth PC (d), for the functional approach.

giving 11 errors (2 false positive and 9 false negative). However, these estimations are optimistic, and we have carried out the double cross-validation strategy explained in Section 2.2, whose estimation has returned 13 errors (2 false positive and 11 false negative).

In other papers, where functional principal component analysis is included as part of the discrimination analysis, such as Hall et al. (2001) or Ramsay and Silverman (2002, Chapter 8), they do not select the components, but they consider the first J components, where J is chosen by cross-validation. Although there is no guarantee that the first components will be the most discriminant, we have also considered this strategy to show the comparison with our selecting strategy.

We have used double cross-validation. First, each observation (bone) is deleted in turn (external cross-validation). For the remaining samples, we carry out the PCA and select the first J components that give the smallest error rate by leave-one-out cross-validation (internal cross-validation) with linear discrimination for the remaining samples. Using the scores of the first J components, we calculate the linear discriminant classifier with the internal samples. The scores of the initially deleted observation for the first J components can be easily obtained as we have their loadings; it is a simple inner product. We use the previously built linear classifier to predict the class label for the initially deleted observation. This prediction is preserved to produce the (external) leave-one-out estimate of the prediction error, because this process is repeated in turn for each of the observations. Of course, for each observation (external iteration), J can vary. By the double cross-validation strategy with the first J components, a total of 13 errors are obtained (3 false positive and 10 false negative). This is the same number of errors as we have obtained with our selecting strategy, with only two components. However, a look at the chosen Js reveals that J very often corresponds to 15 and 16, a very high number taking into account that the eburnated group is constituted by 16 eburnated femora. Therefore, overlearning should not be discarded when the first J components are considered in this case.

2.4. Comparison of the three approaches

First, note the different distribution of the variability for the three approaches. With the first four PCs with landmarks, 92.76% of the variability is considered. However, 84.62% of the variance is concentrated on only one component, whereas the second component only accounts for 3.26% of the variability; therefore, it is not very representative. Predictably, the

landmark approach cannot collect fine details in variation since we only work with 12 points. On the contrary, with the set approach the variability is extremely partitioned: 44 components are needed in order to capture 90.39% of the variability. Interpreting 44 components is not an easy issue. On the other hand, in an intermediate position between the other two approaches, we find the functional PCA, which gives a variability decomposition which is not so extreme: we can obtain more details on variability decomposition than with landmarks, without being so extremely decomposed as in the set approach.

As regards the discriminant problem, obviously (as less information is considered), the classification results for landmarks are the worst. Similar results are obtained for the set and functional approach (15 versus 13 errors, respectively). However, there are other procedures for carrying out functional data classification that will improve the results with respect to the set approach results.

3. ICA in functional data classification

3.1. Curve discrimination

Different alternatives have been proposed for the curve discrimination problem, although mainly for univariate functions. Two of the first methods were a regularized version of LDA called penalized discriminant analysis (PDA) proposed by Hastie et al. (1995), and a generalized linear regression approach proposed by Marx and Eilers (1999). More recently, some non-parametric alternatives have been proposed, such as the kernel one by Ferraty and Vieu (2003), the k-NN one by Burba et al. (2009) or the local linear one by Barrientos-Marin et al. (2010). Biau et al. (2005) also studied k-nearest neighbor classifiers for functional data. Other recent advances in functional data classification appear in the following papers, and all involve some type of preprocessing (sometimes implicit) of the functional data. Rossi and Conan-Guez (2005) suggested the use of neural networks for nonlinear regression and classification of functional data; the use of neural networks was also considered by Ferré and Villa (2006), who studied a preprocessing approach in which functional data are described via a projection on an optimal basis (in the sense of the inverse regression approach), and subsequently submitted to a neural network for further processing; James and Silverman (2005) studied a non-linear regression model for functional descriptors; Rossi and Villa (2006) and Li and Yu (2008) used Support Vector Machines (SVM) for functional data classification. Finally, Epifanio (2008) proposed the use of several shape descriptors. One of these descriptors were the coefficients of independent component analysis (ICA) components. In Epifanio (2008), these descriptors were compared with classical and the most recent advances in univariate functional data classification in three different problems (an artificial problem, a speech recognition problem and a biomechanical application). The first two problems were considered in Ferraty and Vieu (2003), where they also performed a wide-ranging comparative study. In Epifanio (2008), the proposed descriptors were compared with the methodology proposed in Hastie et al. (1995), in Ferraty and Vieu (2003) including the MPLSR method in its semi-metric and PCA, in Rossi and Conan-Guez (2005), in Ferré and Villa (2006), in Rossi and Villa (2006). As Li and Yu (2008) use the same example (a subproblem of the speech recognition problem) as Rossi and Villa (2006), we can also compare the results in Epifanio (2008) with those of Li and Yu (2008). The descriptors proposed in Epifanio (2008) gave results better than or similar to those obtained using the previous techniques (see Epifanio, 2008, for details).

3.2. Coefficients of independent component analysis components

Coefficients of independent component analysis (ICA) components can be computed easily, and provide better than or similar results to those from existing techniques (Epifanio, 2008). Furthermore, they can be extended easily to the multivariate functional case, as we will explain in this section. Although Epifanio (2008) can be seen for details, a brief summary is given here.

Assume that we observe n linear mixtures $x_1(t), \ldots, x_n(t)$ of n independent components $s_j(t)$,

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t), \quad \text{for all } i. \tag{6}$$

In practice, we have discretized curves ($\mathbf{x}_i = \{x_i(t_k); k = 1, \ldots, m\}$), therefore we can consider the $m \times n$ data matrix $\mathbf{X} = \{x_i(t_k)\}$ to be a linear combination of independent components, i.e. $\mathbf{X} = \mathbf{SA}$, where columns of S contain the independent components and $\mathbf{A}$ is a linear mixing matrix. ICA attempts to "un-mix" the data by estimating an un-mixing matrix $\mathbf{W}$ where $\mathbf{XW} = \mathbf{S}$. Under this generative model, the measured "signals" in $\mathbf{X}$ will tend to be "more Gaussian" than the source components (in $\mathbf{S}$) due to the Central Limit Theorem. Thus, in order to extract the independent components or sources, we search for an un-mixing matrix $\mathbf{W}$ that maximizes the non-Gaussianity of the sources.

We compute ICA for functions in the training set. The coefficients in this base ($\mathbf{S}$) can be easily obtained by least squares fitting (Ramsay and Silverman, 2005). If $\mathbf{y} = \{y(t_k)\}_{k=1}^{m}$ is a discretized function, its coefficients are: $(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'y$. These coefficients constitute the feature vector used for the classification step. We assume that all functions are observed at the same points, otherwise we can always fit a basis and estimate the functions at the required points.

Before the application of the ICA algorithm, it is useful to reduce the dimension of the data previously by principal component analysis (PCA) (for details, see Hyvärinen et al., 2000, Section 5), thus reducing noise and preventing overlearning
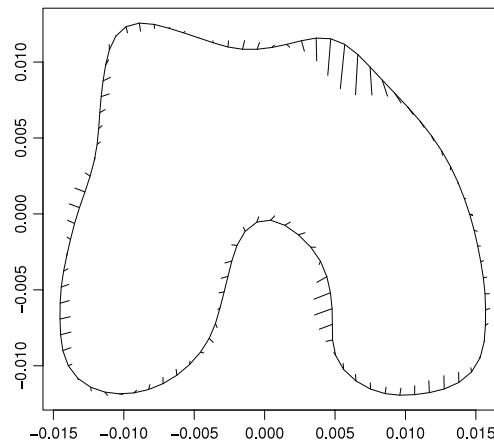
Fig. 5. Mean shape of bones, and $\alpha(t)$ based on ICA.

(Hyvärinen et al., 2001, Section 13.2). Therefore we compute the PCA first, retaining a certain number of components, and then estimate the same number of independent components as the PCA reduced dimension. The FastICA algorithm (which includes the PCA computation in the software available for MatLab and R: http://www.cis.hut.fi/projects/ica/fastica/, with the default parameters, is used for obtaining ICA (Hyvärinen, 1999).

When having multivariate functional data, we can concatenate observations of the functions into a single long vector, as done for computing bivariate functional PCA (Ramsay and Silverman, 2002). Then, the coefficients in ICA base can be used in a classical linear discriminant analysis.

In Epifanio (2008) and Epifanio and Ventura (2008), the components were not selected, but the first J components were considered, where J was chosen by cross-validation, such as in Hall et al. (2001) and Ramsay and Silverman (2002) with functional PCA. Here, besides extending the methodology proposed in Epifanio (2008) to the multivariate functional case, we also propose to select the components in a similar way to that in Section 2 with PCA. Furthermore, we propose to compute a linear discriminant function $\alpha(t)$ based on ICA as done in Ramsay and Silverman (2002, Chapter 8) with PCA. This function $\alpha(t)$ would be the functional counterpart of the linear discriminant or canonical variate (Ripley, 1996, Chapter 3), therefore, $\int_0^1 \alpha(t)x_i(t)\mathrm{d}t$ would return the score or discriminant value of $x_i(t)$ on the discriminant variable.

If we select two ICA functions (I, K) from the ICA base, and apply classical linear discriminant analysis to the ICA coefficients for these two functions, the two coefficients (c_1, c_2) of the linear discriminant are obtained. From them, we can build **a**, a vector of the same length as the number of the ICA basis functions considered, constituted by zeros except for positions (I, K), with values c_1 and c_2, respectively.

The linear discriminant values can be expressed in terms of the ICA coefficients and coefficients of linear discriminants **l**: $\mathbf{l}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{X}$. For a specific individual i: $\mathbf{l}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{x}_i$. At the same time, we can approximate $\int \alpha(t)x_i(t)\mathrm{d}t$ by $\sum_{k=1}^{m} \alpha(t_k)x_i(t_k)$ if we consider the separation between points as one. Therefore, we estimate $\alpha(t)$ at points t_k as $\mathbf{l}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$.

3.3. Bone classification by multivariate functional data discrimination using ICA

We apply the previous methodology to the bones, following the same strategies as for functional PCA in Section 2.3.2. First, we compute ICA with all data. As aforementioned, PCA is calculated before ICA, and only 18 components remain since components associated with eigenvalues of less than 1e-7 are not considered, to avoid singularity of the covariance matrix (as implemented in the FastICA algorithm). For these 18 independent components, the coefficients of 8 components present significant differences between the two groups: 1, 2, 7, 10, 12, 14, 16 and 18. We compute the misclassification rate by leave-one-out, and the best classification with one component is obtained by the first (12 errors, 4 false positive and 8 false negative), while for two components by the first and seventh components (8 errors, 2 false positive and 6 false negative). In Fig. 5, the functional linear discriminant $\alpha(t)$ computed for components 1 and 7, together with the mean shape (in a solid line) are shown (the segments show how the points on the mean shape are perturbed in the direction defined by the discriminant function). The length of the segments shows the zones in which the discriminant score increases most rapidly. In Ramsay and Silverman (2002, Chapter 8), the mode of variability corresponding to a functional linear discriminant $\alpha(t)$ based on the first six principal components of the notch (not the complete bone shape) is displayed. This result is very similar to the one obtained here, if we look at the intercondylar notch.

However, those previous estimations are optimistic, and we have carried out the double cross-validation strategy explained in Section 2.2, whose estimation has returned 11 errors (3 false positive and 8 false negative). As with PCA, we also consider the double cross-validation strategy with the first J components, which gives 12 errors (4 false positive and 8 false negative). This is bigger than the number of errors that we have obtained with our selecting strategy, with only two components. Furthermore, a look at the chosen J's reveals that J very often corresponds to 16, a very high number taking into account that the eburnated group is constituted by 16 eburnated femora.
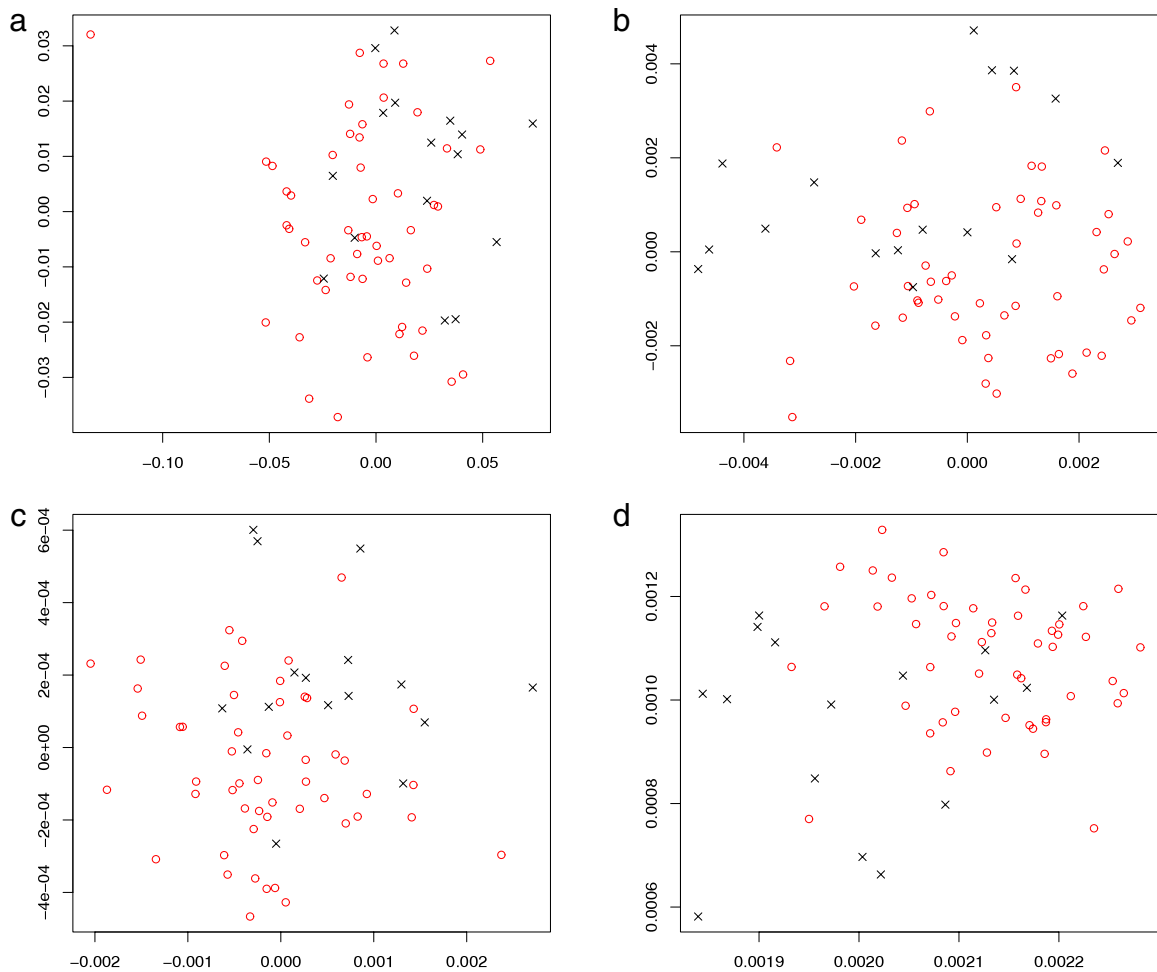
Fig. 6. Scores for PCA components 3 and 7 of landmark approach (a), 2 and 3 of set approach (b), 1 and 6 of functional PCA (c) and 1 and 7 of ICA (d).

It is also very interesting and important to plot the scores of each individual on each component because these scatter plots can reveal interesting features, such as clusters of individuals having a similar behavior on the component in question. For all the approaches, the scores for the two components for which the best result was obtained in each approach is displayed in Fig. 6. The eburnated femora are represented by a cross, while the non-eburnated are represented by a circle. The separation between the two groups is clearer using the ICA approach, where eburnated femora appear mainly in the lower triangular part of the figure.

We now consider the classification obtained by a regularized version of LDA called penalized discriminant analysis (Hastie et al., 1995), using the classical ridge penalty (PDA/Ridge). In the comparative among different methods for curve discrimination (Epifanio, 2008), this method provided an excellent performance. For this method, the shrinkage penalty coefficient (λ) needs to be established. We have considered a double cross-validation for the estimation of the error rate. Several values for λ (1e-04, 5e-05, 1e-05, 5e-06, 1e-06 and 5e-07) have been proposed and used in the *fda* function of the *R* library *mda* (Hastie and Tibshirani, 2006). These proposed values for λ have been selected by computing the performance by leave-one-out for a bigger set of values of λ, where the best result was obtained for $\lambda = $ 1e-5 (the proposed values for λ are around this value): 10 errors (3 false positive and 7 false negative). However, as before, this is an optimistic estimation since λ is selected with the training data. For the double cross-validation strategy, each observation is removed in turn. For each proposed λ, we perform a penalized discriminant analysis leaving-one-out of the remaining samples, and establishing which λ gives the best performance. For this λ, we predict the class of the observation initially removed, using the remaining samples. This prediction is preserved to produce the (external) leave-one-out estimate of the prediction error, because this process is repeated in turn for each of the observations. Of course, for each observation (external iteration), different λs can be selected. By the double cross-validation strategy, the number of errors is 15 (7 false positive and 8 false negative).

Analogously, we have used the nonparametric curve discrimination method (NPCD) with the semi-metric based on functional principal component analysis (FPCA) and multivariate partial least-squares regression (MPLSR) introduced by Ferraty and Vieu (2003) (instead of λ, the tuning parameter is now the number of components used in the semi-metric,
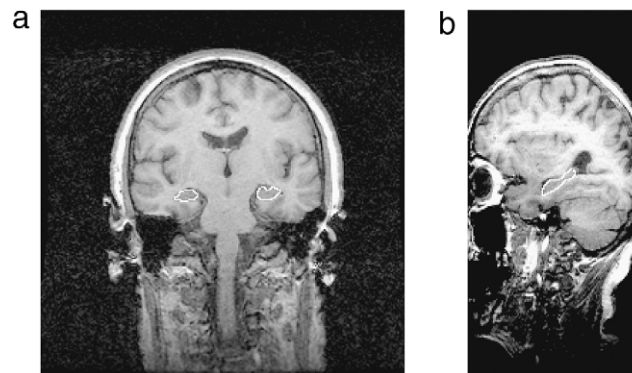
Fig. 7. Hippocampal outlines in a coronal (a) and sagittal (b) slice.

from 1 to 7). The best results using leave-one-out were obtained for 1 component for the first semi-metric (16 errors, 0 false positive and 16 false negative), and 4 factors for the second semi-metric (11 errors, 6 false positive and 5 false negative). Using the double cross-validation strategy, the number of errors is 18 (2 false positive and 16 false negative) for the FPCA semi-metric and 14 (6 false positive and 8 false negative) for the MPLSR semi-metric.

In short, in the functional approach, the best discriminant result is obtained by our methodology, with two ICA components (11 errors), which is better than functional PCA (13 errors) and PDA/Ridge (15 errors) or NPCD/FPCA (18 errors) and NPCD/MPLSR (14 errors). The result with ICA is better than that of the set approach (15 errors) or the landmark approach (16 errors in the optimistic estimate).

4. Hippocampus study in Alzheimer's disease

The early diagnosis of Alzheimer's disease (AD) is a very important issue in our society, since the administration of medicines to subjects who are subtly impaired may render the treatments more effective. Mild cognitive impairment (MCI) is considered as a diagnostic entity within the continuum of cognitive decline towards AD in old age (Grundman et al., 2004; Petersen, 2004). Longitudinal studies indicate a direct relation between the hippocampal volume decrease in and cognitive decline (Jack et al., 1999; Mungas et al., 2001). However, volumetric measurements are simplistic characteristics and structural changes at specific locations cannot be reflected in them. If morphological changes could be established, then this should enable researchers to gain an increased understanding of the condition. This is the reason why nowadays shape analysis is of an enormous importance in neuroimaging circles (Styner et al., 2004).

In order to understand the way in which hippocampi differ among three different groups (controls, patients with MCI, and patients with early AD), we have their magnetic resonance (MR) scans, which will be transformed into multivariate functional data, as explained in following section. These multivariate functional data will be used in a functional discriminant analysis. We will apply the methodology presented in the previous section, with some small modifications.

4.1. Brain MR scans processing

28 individuals were analyzed in this study: 12 controls (5 males and 7 females, with mean age 70.17 and standard deviation 3.43), 6 patients with MCI (2 males and 4 females, with mean age 75.50 and standard deviation 3.33), and 10 patients with early AD (1 male and 9 females, with mean age 71.50 and standard deviation 4.35). All the subjects were recruited from the Neurology Service at La Magdalena Hospital (Castelló, Spain) and the Neuropsychology Service at the Universitat Jaume I. All experimental procedures complied with the guidelines of the ethical research committee at the Universitat Jaume I. Written informed consent was obtained from every individual or their appropriate proxy prior to participation. Selection for the participant group was made after careful neurological and neuropsychological assessment. The neuropsychological test battery involved Digit Span, Similarities, Vocabulary, and Block Design of the WAIS-III; Luria's Watches test, and Poppelreuter's Overlapping Figure test. MR scans were carried out with a 1.5T General Electric system. A whole brain high resolution 3D-Gradient Echo (FSPGR) T1-weighted anatomical reference scan was acquired (TE 4.2 ms, TR 11.3 ms, FOV 24 cm; matrix $= 256 \times 256 \times 124$, 1.4 mm-thick coronal images).

Hippocampi were traced manually on contiguous coronal slices (or sections) following the guidelines of Watson et al. (1992), and Hasboun et al. (1996). The hippocampus segmentation was done by an expert tracer with MRIcro software, blinded to the clinical data of the study subjects. The segmentation of each hippocampus lasted approximately 40 min. An example of the left and right hippocampal contour (drawn in white) in a coronal view can be seen in Fig. 7(a), while a sagital view of one of the hippocampus can be seen in Fig. 7(b). Each hippocampus is described by around 30 coronal slices.

As hippocampal volume is related to the patient's condition, we describe each slice by its area (note that if these areas are added up, we obtain an amount proportional to the volume). The area for each slice can be estimated as the number of
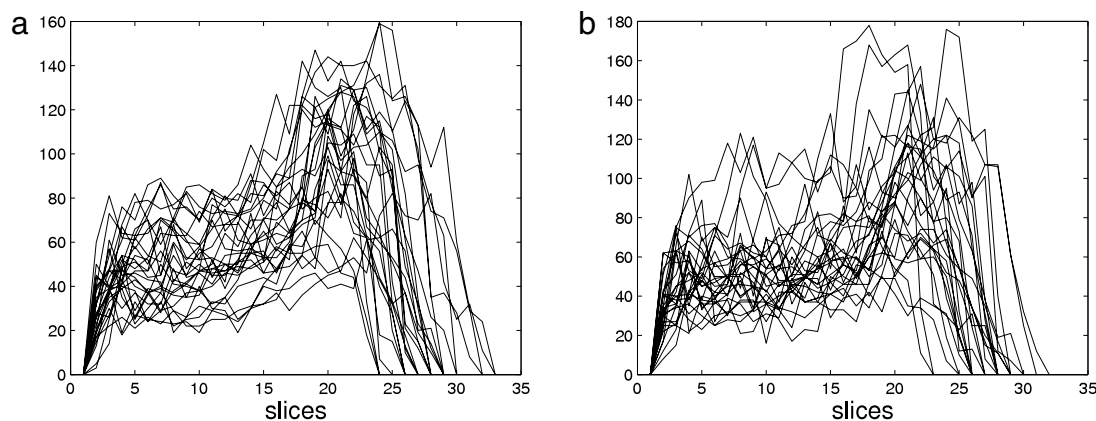
Fig. 8. Right (a) and left (b) hippocampal areas for each slice.

pixels belonging to each segmented hippocampal slice. The area of the left and right hippocampus in each slice is computed. Therefore, for each subject we have two functional data. However, the argument is not time, as is usual, but space: the coronal axis, the slices. Only the left and right hippocampal area in each coronal slice (with 1.4 mm of separation between them) are available. These functions are measured discretely, but in principle they really represent continuous functions, since (obviously) the hippocampus is a continuous structure in the space. Therefore, they are functional data. In Fig. 8, we can see the two functional data for the 28 individuals. As the coronal length of each hippocampus is variable, in order to have a common axis (33 slices), we complete the raw data by adding zeros when the hippocampal surface is finished. Moreover, for all subjects, the first and last slice areas are zero.

4.2. Results of the hippocampus study

First, we consider the right and left hippocampal volumes, which are classical variables in the literature. Volumes are estimated as the sum of the slice areas. When volumes of right and left hippocampi are considered, the number of errors are 5, using linear discriminant analysis and leave-one-out cross-validation. However, if hippocampi are considered separately, 6 and 7 errors are obtained for the left and right hippocampi, respectively.

Second, bivariate functional data compiling slice areas for the left and right hippocampi are considered. Data are smoothed by 31 (nearly the same number of slices due to the low sampling rate) Fourier basis functions. The next step is to perform a registration process in order to take into account the phase variation (some hippocampi only appear in 24 slices). One registration for the left and another one for the right hippocampus function are carried out, applying the function *registerfd* of the package *fda* (Ramsay et al., 2009), using the minimum eigenvalue of a cross-product matrix as the continuous registration criterion and the mean function as the target function (Ramsay and Silverman, 2005). Default parameters are used for this function, except for the number of iterations, which is increased in order to satisfy the criterion for convergence.

We now use the methodology presented in Section 3.2 with some small changes due to the nature of the problem. In this problem the number of samples (slices) practically equals the dimension of the data (28). Although we could compute as many independent components as signals, this results in overlearned components having single spikes (Särelä and Vigário, 2003), and these components are not interpretable. As explained in Särelä and Vigário (2003), one solution for circumventing overlearning could be to acquire more samples, but in this case it is not possible to obtain more slices (this would increase the already long acquisition time of MR brain scans). Another solution is reducing dimension: the number of free parameters is $n^2/2$, so in this problem we compute a maximum of 7 components ($n < \sqrt{2 \cdot 30}$), taking into account that each hippocampus is described by around 30 slices (samples). So, the first modification is that we compute the independent components varying their number from 1 to 7. For each case, we consider the components for which the difference among the 3 groups is significant. However, in this problem the t-test cannot be used as before because we have 3 groups, not 2, so we use the Kruskal–Wallis test.

As our objective is the way in which the hippocampi differ among the three groups, we show the combination for which the best discriminant results are obtained, by leave-one-out. The error estimate obtained in this way will be optimistic, but we want to investigate the shape variation among the groups, obtaining the linear discriminant functions.

The number of misclassifications for the left hippocampi is 2, with 3 significant components (1, 2 and 5) from the 5 computed, whereas it is 7 for the right hippocampi, with 2 components, from the 2 components computed, which are significant. Note that in Epifanio and Ventura (2008), when we used all the components without selection, the best result was 3 errors, obtained with 5 components for the left hippocampi. Fig. 9(a) and (b) display the mode of variability corresponding to the resulting $\alpha(t)$s, for the left and right hippocampi, respectively, with the vertical lines. The solid curve is the mean. The
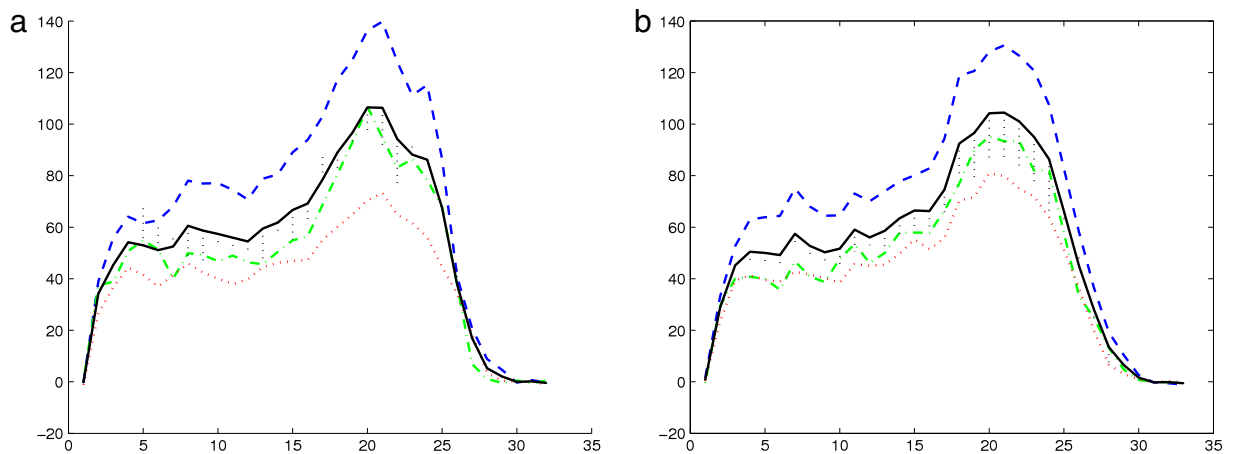
Fig. 9. The mode of variability corresponding to $\alpha(t)s$, for the left (a) and right (b) hippocampi. See the text for details.
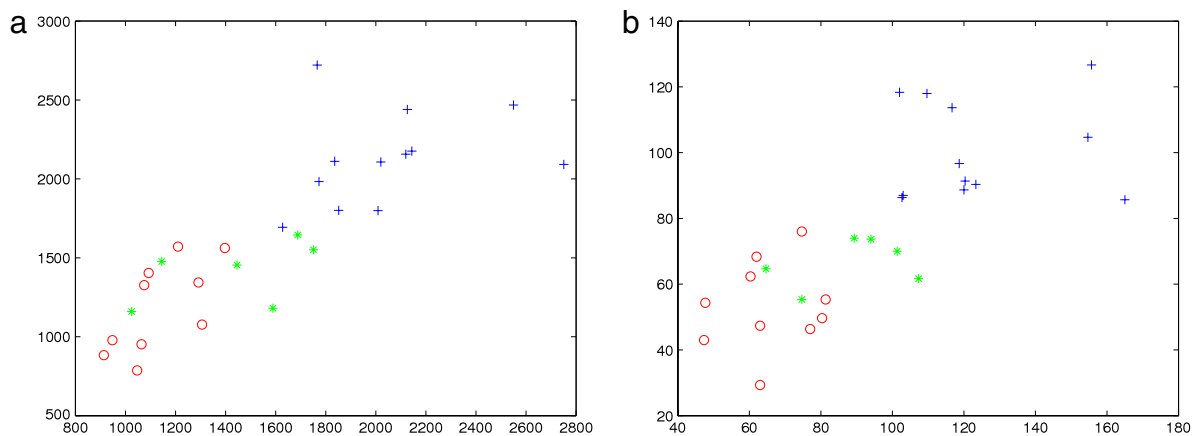

Fig. 10. Left vs. right hippocampal volumes (a), and mean of the areas of the determined hippocampal head slices (b). Crosses, stars and circles indicate the controls, patients with MCI, and patients with early AD, respectively.

dashed, dash–dotted and dotted curves represent the mean of the controls, patients with MCI, and patients with early AD, respectively. The first linear discriminant explains 96.82 % and 91.86% of the variance between groups, for the left and right hippocampi, respectively.

Head, body and tail are the three parts that make up a hippocampus (Hasboun et al., 1996). Both for the right and left hippocampus, the zone where the linear discriminant functions are bigger in absolute value corresponds to the head of the hippocampi. In other studies, with other methodologies, the same conclusion was reached (Wang et al., 2003). In particular, the point where $\alpha(t)s$ takes its maximum absolute value for the left hippocampus is 22, while it is 24 for the right hippocampus. We can find the left and right slices of each subject corresponding to those maximum values, inverting the corresponding warping function of the registration process and rounding to the nearest integer. Besides the (left and right) slice obtained for each individual with this process, we also consider the previous and subsequent slice to the determined slice, so as not to base the following analysis only on one single slice. In short, three slices for the right and three slices for the left hippocampi are considered for each subject. The averages of the areas of these three slices are shown in Fig. 10(b), together with the volumes for the right and left hippocampi (Fig. 10(a)). In Fig. 10(b) is easier to discriminate between groups. The number of misclassifications with the mean hippocampal head areas is 3 by leave-one-out.

Instead of considering simply the area of the three slices determined, we can consider their outlines. We parameterize the outlines of each of the three determined slices by arc length with 45 points. The different slices are translated to the origin in such a way that their centroids coincide with the origin. The tracing begins counterclockwise in the easternmost outline point in the same row as the centroid, using the function *bwtraceboundary* of the image toolbox of MatLab, as was done before with the bones. A polygonal basis (Ramsay et al., 2009) with fifteen functions is used to represent these outlines. The averages of the three slices considered per individual are calculated for the left and right hippocampi. Therefore, we have two pairs of functions $\{X(t), Y(t)\}$ (which represent the mean outlines of the three determined slices, $X(t)$ gives the X coordinates, while $Y(t)$ the coordinates in the Y axis) for each individual, one pair for the right and another pair for the
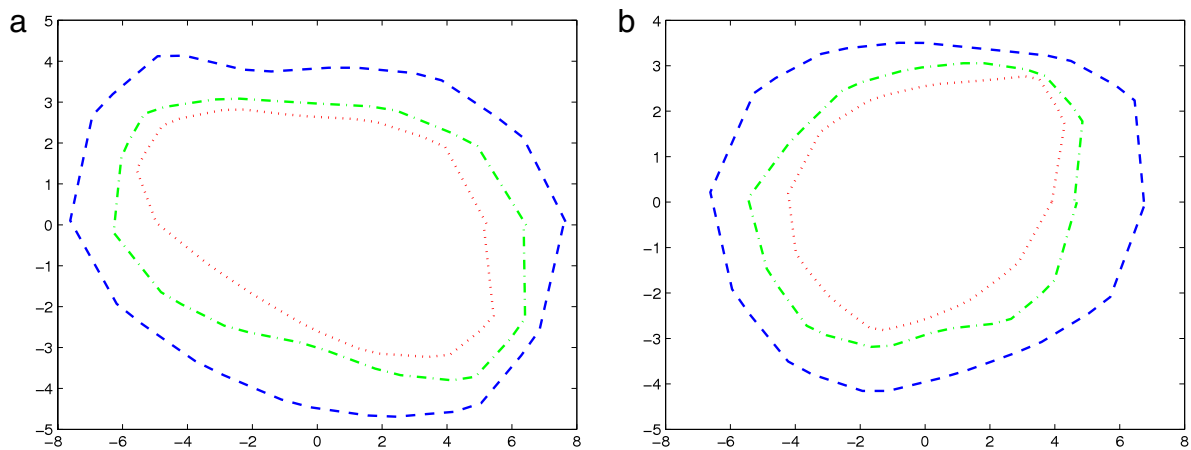
Fig. 11. Averages for the left (a) and right (b) slices of the determined hippocampal head slices.

left hippocampus, i.e. a total of four functional data per individual. In Fig. 11 the averages of these functions are shown. The dashed, dash–dotted and dotted curves represent the mean of the controls, patients with MCI, and patients with early AD, respectively.

Using these four functions jointly, we have deleted one individual each time and computed ICA without that observation. The linear discriminant scores obtained with the 27 individuals are used to predict the class of the observation deleted. With this leave-one-out strategy and one independent component, only 2 misclassifications are achieved, which are very promising results. Remember that we selected the slices of the hippocampal head with the 28 individuals, and therefore the estimated error will be optimistic, but this application is simply an illustration of the methodology. Note that in Epifanio and Ventura (2008), without selecting the components, we also obtained 2 errors but using three components instead of one. With the set approach (concatenating the images for the right and left hippocampi obtained with the mean outlines), 4 errors are obtained with two components, the most discriminant ones as explained in Section 2.2 (with only one component, 4 errors are also obtained). With functional PCA and two components (the most discriminant ones in each step of the double cross-validation, as explained in Section 2.3.2) 8 errors are obtained (10 errors with one component). Therefore, the advantage of using ICA is clear.

5. Conclusions and discussion

In this study we have revealed the role that FDA can play in image analysis, where despite many descriptors being functions, FDA has hardly been applied. In particular, we have focused our attention on shape analysis, where the use of FDA in the functional approach has shown its superiority over other approaches, such as the landmarks or the set theory approach, in two different problems (PCA and discriminant analysis) in a well-known database of bone outlines (see Section 2.4).

Furthermore, we have dealt with a problem that has hardly ever been considered in the literature: the multivariate functional discrimination (most of the existing literature on functional classification considers univariate curves datasets). We have also proposed a discriminant function based on ICA, and the classification results obtained with this methodology are very promising.

Unlike other papers in the literature where functional PCA is included as part of the discrimination analysis, we have proposed selecting the components. This has also been proposed with ICA. Our option has been to consider all the components and select the significant ones. However, in the hippocampus study, the number of samples practically equals the dimension of the data, and we had to make a small modification in order to avoid overlearned components. The problem of how many components should be estimated is an open question in ICA (Hyvärinen et al., 2001). This a point for future study, and maybe an order could be established such as in Cheung and Xu (1999). Another point to study is improving the smoothness of the functions by using a roughness penalty.

We have also applied FDA in the analysis of MR scans in order to study the hippocampal differences among the subjects in three groups: controls, MCI, and patients with early Alzheimer's disease. The database is quite small for obtaining valid medical conclusions, although the methodology could be used without modification with a larger database. This is a novel application of FDA in image analysis, where we use a spatial argument for the functional data instead of the temporal argument commonly used in FDA.

In the application, we have seen that the head was the most discriminative part. This point is very interesting, since if segmentation was reduced only to the hippocampal head, the segmentation time would be shorter. Furthermore, it is easier to implement an automatic segmentation for the hippocampal head only, which will decrease that time even more, and will eliminate variability due to the subjectivity of the manual tracer. (Remember that the total time for the manual segmentation of one hippocampus was approximately 40 min.).

Some additional points to study are as follows: first, from the statistical point of view, the use of ICA in other statistical problems such as functional logistic regression or visualization (Hyndman and Shang, 2010), or the use of ICA in the construction of a semi-metric for use with non-parametric techniques (Ferraty and Vieu, 2006) (PCA-type semi-metrics could be replaced by ICA-type semi-metrics by just considering the ICA expansion instead of the PCA-based expansion. This is a very interesting field of study, since compared with PCA, ICA allows better observation of the underlying structure of the data. PCA is a purely second-order statistical method, whereas ICA requires the use of higher-order statistics; therefore, ICA can be seen as an extension to PCA.). Second, from the image analysis point of view, is the application of FDA in other problems of shape analysis such as the definition of confidence and quantile sets (Simó et al., 2004), or its use when the closed contour of a figure is not always available, such as in Domingo et al. (2005), maybe using a discontinuous function. Third, FDA can be exploited in other fields of image analysis besides shape analysis, such as texture analysis (Epifanio et al., 2009). Finally, in order to introduce the methodology more easily, we have restricted the analysis to two dimensional outlines, but FDA can be used for surfaces (multidimensional functions with two arguments). In fact, in the future, we are going to work in three dimensions in the hippocampal analysis. The methodology presented can be extended for functions with two (or more) arguments.

Acknowledgements

References

Ambroise, C., McLachlan, G.J., 2002. Selection bias in gene extraction on the basis of microarray gene-expression data. Proceedings of the National Academy of Sciences 99 (10), 6562–6566.

Barrientos-Marin, J., Ferraty, F., Vieu, P., 2010. Locally modelled regression and functional data. Journal of Nonparametric Statistics 22 (5), 617–632.

Berrendero, J.R., Justel, A., Svarc, M., 2011. Principal components for multivariate functional data. Computational Statistics and Data Analysis doi:10.1016/j.csda.2011.03.011.

Biau, G., Bunea, F., Wegkamp, M., 2005. Functional classification in Hilbert spaces. IEEE Transactions on Information Theory 51, 2163–2172.

Burba, F., Ferraty, F., Vieu, P., 2009. k-Nearest Neighbour method in functional nonparametric regression. Journal of Nonparametric Statistics 21 (4), 453–469.

Cheung, Y., Xu, L., 1999. MSE reconstruction criterion for independent component odering in ICA time series analysis. In: Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing. pp. 793–797.

Delicado, P., 2011. Dimensionality reduction when data are density functions. Computational Statistics and Data Analysis 55 (1), 401–420.

Domingo, J., Nacher, B., de Ves, E., Alcantara, E., Díaz, E., Ayala, G., Page, A., 2005. Quantifying mean shape and variability of footprints using mean sets. In: Proceedings of the 7th International Symposium on Mathematical Morphology. Springer, pp. 455–464.

Dryden, I., 2007. Shapes: Statistical shape analysis. http://www.maths.nott.ac.uk/~ild/shapes.

Dryden, I.L., Mardia, K.V., 1998. Statistical Shape Analysis. Wiley, Chichester.

Epifanio, I., 2008. Shape descriptors for classification of functional data. Technometrics 50 (3), 284–294.

Epifanio, I., Domingo, J., Ayala, G., 2009. Texture classification by functional analysis of size distributions. In: Proceedings of the 11th IASTED International Conference on Signal and Image Processing. pp. 172–177.

Epifanio, I., Ventura, N., 2008. Multivariate functional data discrimination using ICA: analysis of hippocampal differences in Alzheimer's disease. In: Contributions to Statistics. Springer, pp. 157–163. (Chapter 25).

Ferraty, F., Romain, Y., 2011. The Oxford Handbook of Functional Data Analysis. Oxford University Press.

Ferraty, F., Vieu, P., 2003. Curves discrimination: a nonparametric functional approach. Computational Statistics and Data Analysis 44, 161–173.

Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer.

Ferré, L., Villa, N., 2006. Multilayer perceptron with functional inputs: an inverse regression approach. Scandinavian Journal of Statistics 33 (4), 807–823.

González, R., Woods, R., Eddins, S., 2004. Digital Image Processing Using MATLAB. Prentice Hall.

González-Manteiga, W., Vieu, P., 2007. Statistics for functional data. Computational Statistics and Data Analysis 51 (10), 4788–4792.

Gower, J.C., 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325–338.

Grundman, M., Petersen, R.C., Ferris, S.H., Thomas, R.G., Aisen, P.S., Bennett, D.A., Foster, N.L., Jack, Jr., C.R., Galasko, D.R., Doody, R., Kaye, J., Sano, M., Mohs, R., Gauthier, S., Kim, H.T., Jin, S., Schultz, A.N., Schafer, K., Mulnard, R., van Dyck, C.H., Mintzer, J., Zamrini, E.Y., Cahn-Weiner, D., Thal, L.J., 2004. Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. Archives of Neurology 61, 59–66.

Hall, P., Poskitt, D., Presnell, B., 2001. A functional data-analytic approach to signal discrimination. Technometrics 43, 1–9.

Hasboun, D., Chantôme, M., Zouaoui, A., Sahel, M., Deladoeuille, M., Sourour, N., Duyme, M., Baulac, M., Marsault, C., Dormont, D., 1996. MR determination of hippocampal volume: comparison of three methods. American Journal of Neuroradiology 17, 1091–1098.

Hastie, T., Buja, A., Tibshirani, R., 1995. Penalized discriminant analysis. Annals of Statistics 23, 73–102.

Hastie, T., Tibshirani, R., 2006. mda: Mixture and flexible discriminant analysis. R port by Leisch, F., Hornik, K. and Ripley, B.D.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Data Mining, Inference and Prediction, 2nd ed.. Springer-Verlag.

Horgan, G.W., 2000. Principal component analysis of random particles. Journal of Mathematical Imaging and Vision 12, 169–175.

Hyndman, R.J., Shang, H.L., 2010. Rainbow plots, bagplots and boxplots for functional data. Journal of Computational and Graphical Statistics 19 (1), 29–45.

Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for indepent component analysis. IEEE Transactions on Neural Networks 10 (3), 626–634. http://www.cis.hut.fi/projects/ica/fastica/.

Hyvärinen, A., Karhunen, J., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Networks 13, 411–430.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley, New York.

Jack, C.R., Petersen, R.C., Xu, Y.C., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Waring, S.C., Tangalos, E.G., Kokmen, E., 1999. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. Neurology 52, 1397–1403.

James, G.M., Silverman, B.W., 2005. Functional adaptive model estimation. Journal of the American Statistical Association 100, 565–576.

Jolliffe, I.T., 2002. Principal Component Analysis, 2nd ed.. Springer.

Kindratenko, V.V., 2003. On using functions to describe the shape. Journal of Mathematical Imaging and Vision 18, 225–245.

Li, B., Yu, Q., 2008. Classification of functional data: a segmentation approach. Computational Statistics and Data Analysis 52 (10), 4790–4800.

Li, P.L., Chiou, J.M., 2011. Identifying cluster number for subspace projected functional data clustering. Computational Statistics and Data Analysis 55 (6), 2090–2103.

López-Pintado, S., Romo, J., 2011. A half-region depth for functional data. Computational Statistics and Data Analysis 55 (4), 1679–1695.

Marx, B., Eilers, P., 1999. Generalized linear regression on sampled signals and curves: a P-spline approach. Technometrics 41, 1–13.

Mungas, D., Jagust, W.J., Reed, B.R., Kramer, J.H., Weiner, M.W., Schuff, N., Norman, D., Mack, W.J., Willis, L., Chui, H.C., 2001. MRI predictors of cognition in subcortical ischemic vascular disease and Alzheimer's disease. Neurology 57, 2229–2235.

Nettel-Aguirre, A., 2008. Nuclei shape analysis, a statistical approach. Image Analysis and Stereology 27, 1–10.

Petersen, R.C., 2004. Mild cognitive impairment as a diagnostic entity. Journal of Internal Medicine 256, 183–194.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, ISBN: 3-900051-07-0. http://www.R-project.org.

Ramsay, J.O., Hooker, G., Graves, S., 2009. Functional Data Analysis with R and MATLAB. Springer.

Ramsay, J.O., Silverman, B.W., 2002. Applied Functional Data Analysis. Springer.

Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, 2nd ed.. Springer.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.

Rossi, F., Conan-Guez, B., 2005. Functional multi-layer perceptron: a nonlinear tool for functional data analysis. Neural Networks 18 (1), 45–60.

Rossi, F., Villa, N., 2006. Support vector machine for functional data classification. Neurocomputing 69 (7–9), 730–742.

Särelä, J., Vigário, R., 2003. Overlearning in marginal distribution-based ICA: analysis and solutions. Journal of Machine Learning Research 4, 1447–1469.

Shepstone, L., Rogers, J., Kirwan, J., Silverman, B., 1999. The shape of the distal femur: a palaeopathological comparison of eburnated and non-eburnated femora. Annals of the Rheumatic Diseases 58, 72–78.

Shepstone, L., Rogers, J., Kirwan, J., Silverman, B., 2000. Distribution of distal femoral osteophytes in a human skeletal population. Annals of the Rheumatic Diseases 59, 513–520.

Simó, A., de Ves, E., Ayala, G., 2004. Resuming shapes with applications. Journal of Mathematical Imaging and Vision 20, 209–222.

Soille, P., 2003. Morphological Image Analysis. Principles and Applications, 2nd ed.. Springer-Verlag.

Stoyan, D., Stoyan, H., 1994. Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics. Wiley.

Styner, M., Lieberman, J.A., Pantazis, D., Gerig, G., 2004. Boundary and medial shape analysis of the hippocampus in schizophrenia. Medical Image Analysis Journal 8 (3), 197–203.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S-plus. Springer.

Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernanskya, J.G., 2003. Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. NeuroImage 20, 667–682.

Watson, C., Andermann, F., Gloor, P., Jones-Gotman, M., Peters, T., Evans, A., Olivier, A., Melanson, D., Leroux, G., 1992. Anatomic basis of amygdaloid and hippocampal volume measurement by magnetic resonance imaging. Neurology 42 (9), 1743–1750.