UNIVERSIDAD CARLOS III DE MADRID

working papers

UC3M Working Papers
Statistics and Econometrics
17-05
ISSN 2387-0303
Abril 2017

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-48

KERNEL DEPTH FUNCTIONS FOR FUNCTIONAL DATA

Nicolás Hernández[a], Alberto Muñoz[a]

Abstract

In the last years the concept of data depth has been increasingly used in Statistics as a center-outward ordering of sample points in multivariate data sets. Recently data depth has been extended to functional data. In this paper we propose new intrinsic functional data depths based on the representation of functional data on Reproducing Kernel Hilbert Spaces, and test its performance against a number of well known alternatives in the problem of functional outlier detection.

Keywords: Kernel depth; Functional Data Analysis; Reproducing Kernel Hilbert Spaces; Outlier detection.

[a] Department of Statistics, Universidad Carlos III de Madrid.

Acknowledgements: the authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness ECO2015-66593-P.

Kernel depth functions for functional data

Nicolás Hernández & Alberto Muñoz

Departamento de Estadística

Universidad Carlos III de Madrid, Spain

May 29, 2017

Abstract

In the last years the concept of data depth has been increasingly used in Statistics as a center-outward ordering of sample points in multivariate data sets. Recently data depth has been extended to functional data. In this paper we propose new intrinsic functional data depths based on the representation of functional data on Reproducing Kernel Hilbert Spaces, and test its performance against a number of well known alternatives in the problem of functional outlier detection.

Keywords: Kernel depth · Functional Data Analysis · Reproducing Kernel Hilbert Spaces · Outlier detection.

1 Introduction

In functional analysis functions are considered as points in a vector space. Being functions infinite-dimensional objects, functional data analysis (FDA) techniques must deal with this property. A functional datum (or curve) is defined by $f_n = \{(\mathbf{x}_i; y_i) \in X \times Y\}_{i=1}^n$, where X is the input variables space, and usually $Y \in \mathbb{R}$. The first step to deal with this type of data is to obtain a function $f : X \to Y$ to approximate f_n and then apply any mathematical procedure capable of analyzing this kind of objects. In this sense, a core idea in FDA is to consider functional data as points in a function space, as a previous step to the projection of such functions onto a finite dimensional Euclidean space. This process necessarily involves obtaining new representations for functional data. In some cases this approach can be achieved by using orthogonal basis of functions $B = \{\phi_i\}, i \in I$ (where I is some index set) of some functional subspace included in $L^2(X)$. The usual basis functions are Fourier, Wavelets or Spline basis, see Wahba (1990).

Our proposal, which is described in detail in Section 3, involves the use of kernel methods to represent each functional datum. We will adopt the procedure in Muñoz and González (2010), that considers each curve as a multivariate point in a functional space and then project these points into a Reproducing Kernel Hilbert Space (RKHS) using regularization theory techniques.

In this work we focus on the task of inducing a total order on the data set associated to depth. Consider a data set $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\} \in \mathbb{R}^d$. When $d = 1$ the degree of centrality of a given point $\mathbf{x}_i$ with respect to a probability distribution can be defined by ranking all the distances from each point to the median. In the multivariate case ($d \geq 2$), we first define the central (deepest) point of the distribution/data set and the degree of centrality is given by ranking the distances from each to the deepest one using some metric. A function that implements that mapping is called in the literature 'depth function'. Thus, depth functions compute how deep is a point with respect to a distribution/data set (Liu et al., 1999, Zuo and Serfling, 2000), defining the degree of 'centrality' or 'outlyingness' of a point in a multivariate data set given an underlying distribution.

Depth can be extended to functional data in several ways. López-Pintado and Romo (2009) introduced the concept of 'bands' to compute the deepest curve. For a sample of n curves, the modified band depth method consider 'bands' defined for combinations of $2, 3, \ldots$ up to n curves, and accounts for the proportion of 'x' axis coordinate that a curve c_l is contained in the band (depth index). Hence, the depth of c_l is defined as the average of the depth index for all the possible bands. The deepest curve is the curve with the maximum depth. The half-region depth and the modified half-region depth (López-Pintado and Romo, 2011) allow to order a functional data set taking into account the hypograph and epigraph of a functional datum. The Fraiman and Muniz depth or Integrated depth measures the conditional quantile on all points. In particular when working with curves it measures how long a curve remains in the middle of a sample of curves, (Fraiman and Muniz, 2001). In this contributions the authors define functional data depths measures from the original plain representation $\{(\mathbf{x}_i, y_i)\}$ of the data points (i.e. curves). Our objective is to define a new functional depth measure, but as it was stated before, the core idea is to work with the functional representation of the functional data.

The rest of the paper is organized as follows: in Section 2 we present a review of several multivariate depth measures and its properties, and an extension to the functional framework as well. A new methodological proposal of a functional depth is described in Section 3, as well as the proofs of its properties. In Section 4 the results of the empirical work are shown for a set of simulated and real functional data, and we conclude in Section 5 with some final considerations and future research lines.

2 Review of depth measures

Up to now we have introduced the concept of depth. Formally in Zuo and Serfling (2000) the authors formally define a statistical depth function as follows:

Definition 2.1. Statistical depth function [Zuo and Serfling (2000)]. Let the mapping $D(\cdot;\cdot)$: $\mathbb{R}^d \times \mathcal{F} \to \mathbb{R}$ be bounded, non-negative and satisfies properties 1-4, where $\mathcal{F}$ class of distributions on the Borel sets of $\mathbb{R}^d$ and by $F_{\mathbf{x}}$ the distribution of $\mathbf{x}$:

Property 1. *Affine invariance.* This property states that the depth of a given point does not change if an affine transformation is applied. $D(\mathbf{x}; F_{\mathbf{x}}) = D(A\mathbf{x} + b; F_{A\mathbf{x}+b})$.

Property 2. *Maximality at center.* The depth function should attain the maximum value at the center of the distribution (uniquely defined). $D(\mathbf{x}_0; F) = \sup_{\mathbf{x}} D(\mathbf{x}; F)$, for any $\mathbf{x}_0 \in \mathbb{R}^d$ that is the center of F.

Property 3. *Monotonicity relative to the deepest point.* As any point $\mathbf{x} \in \mathbb{R}^d$ turns away from the deepest point, the depth of $\mathbf{x}$ should decrease monotonically. $D(\mathbf{x}; F) \leq D(\alpha\mathbf{x} + (1-\alpha)\mathbf{x}_0; F)$, for any $\mathbf{x}_0 \in \mathbb{R}^d$ that is the center of F and $\alpha \in [0, 1]$.

Property 4. *Vanishing at infinity.* The value of the depth function at any point $\mathbf{x} \in \mathbb{R}^d$ should tend to zero as the $\|\mathbf{x}\|$ goes to infinity. For each F, $D(\mathbf{x}; F) \to 0$, as $\|\mathbf{x}\| \to \infty$

Then $D(\cdot; F)$ is a statistical depth function.

In the following subsection the most widely used measures are presented, namely: i) the Mahalanobis depth, (MhD), ii) the half-space depth, (HD) and iii) the simplicial depth (SD). For a formal and extensive review of different depth measures and its properties, see (Liu et al., 1999) and Zuo and Serfling (2000).

2.1 Multivariate depth measures

Let F be an absolutely continuous probability distribution in $\mathbb{R}^d$, with $d \geq 1$, and $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ a random sample of F, where each $\mathbf{x}_i$ is a column vector $d \times 1$. All the measures give the depth of a given point $\mathbf{x}$ relative to the distribution F.

Definition 2.2. Mahalanobis depth [Mahalanobis (1936)].

$$M_h D(\mathbf{x}; F) = [1 + (\mathbf{x} - \mu_F)\Sigma_F^{-1}(\mathbf{x} - \mu_F)]^{-1},$$

where μ_F and Σ_F are the mean vector and the covariance matrix of the distribution F. To obtain the sample version, μ_F and Σ_F must be substituted by their sample estimators.

Definition 2.3. Half-space (*Tukey*) depth [Tukey (1975)].

$$HD(\mathbf{x}; F) = \inf_H \{P(H) : x \in H\},$$

where H is a closed halfspace in $\mathbb{R}^d$ and $\mathbf{x} \in H$. For the sample version F must be replaced by the empirical distribution F_n. The *Tukey* depth w.r.t a data set considers the minimum number of sample points of a distribution that belongs to one side of a hyperspace (halfspace) through the point $\mathbf{x}_i$.

Definition 2.4. Simplicial depth [Liu et al. (1990)].

$$SD(\mathbf{x}; F) = P_F\{\mathbf{x} \in S[\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}]\},$$

where $S[\mathbf{x}_1, \ldots, \mathbf{x}_{d+1}]$ is a closed simplex of $(d+1)$ random observations of F. The idea is that if we have data in $\mathbb{R}^d$ form all the possible simplices, and the deepest point will be the one that belongs to more simplices. To estimate the sample simplicial depth, an indicator function $\mathbf{I}_{(\cdot)}$ must be introduced to have

$$SD(\mathbf{x}; f_n) = \binom{n}{d+1}^{-1} \sum \mathbf{I}_{(\mathbf{x} \in S[\mathbf{x}_{i1}, \ldots, \mathbf{x}_{id+1}])}.$$

It is relevant to analyze wether these depth measures satisfy the properties 1-4.The *half-space* or *Tukey* depth satisfy the four properties described above. Also the *Mahalanobis* depth satisfy that four properties but only when F is symmetric. With respect to the *Simplicial* depth function, it also satisfies properties from 1-4 but for the case when F is an angularly symmetric[1] distribution, in other cases properties 2 and 3 are not always satisfied see (Zuo and Serfling, 2000). Moreover Serfling (2006) mention other four properties that are desirable but not necessary, and are listed below.

- i *Symmetry.* Let $\mathbf{x}_0$ be the deepest point, if F is symmetric around $\mathbf{x}_0$ then so it is $D(\mathbf{x}; F_\mathbf{x})$.
- ii *Continuity of $D(\mathbf{x}, F)$ as a function of* $\mathbf{x}$, (upper semicontinuity).
- iii *Continuity of $D(\mathbf{x}, F)$ as a function of F.*
- iv *Quasi-concavity as a function of* $\mathbf{x}$. The set $\{\mathbf{x} : D(\mathbf{x}, F) \geq c\}$ is convex for each real c.

[1] X present an angularly symmetric distribution in θ if $\frac{(X-\theta)}{\|X-\theta\|}$ is centrally symmetric in θ, which means $X - \theta \stackrel{d}{=} \theta - X$, see Liu et al. (1990) for further details.

2.2 Notion of functional depth measures

The extended concept of depth to functional data has the same objective: measure the degree of centrality of a point, in this case a curve, with respect to a sample of functional data. In the next subsections we describe the most widely used depth measures for functional data and its properties.

2.2.1 The band depth measure

The *band-depth* measure arose from a graph-based approach as a methodology to find the depth of an element in a given space with respect to a sample of functional data. It can be considered as a functional extension of the idea proposed in the simplicial depth by Liu et al. (1990). To define precisely the band-depth measure is important to introduce some previous concepts.

Given a compact interval I, let $C(I)$ be the space of the real continuous functions. Given $X = \{x_1(t), \ldots, x_n(t)\}$, the graph of a function x, is defined as $G(x) = \{(t, x(t)) : t \in I\}$. Also, the band in $\mathbb{R}^2$ delimited by the curves $x_{i1}, \ldots, x_{ik} \subset X$, is

$$B(x_{i1}, \ldots, x_{ik}) = \{(t, y) : t \in I, \min_{r=1,\ldots,k} x_{ir}(t) \leq y \leq \max_{r=1,\ldots,k} x_{ir}(t)\} \tag{2.1}$$

Then the proportion of bands $B(x_{i1}, \ldots, x_{ik})$ determined by j different curves $x_{i1}, \ldots, x_{ik}$ containing the entire graph of x is

$$BD_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} \mathbf{I}\{G(x) \subseteq B(x_{i1}, \ldots, x_{ik})\}. \tag{2.2}$$

Now the band-depth definition can given as the sum on J of eq. 2.2.

Definition 2.5. The band-depth [López-Pintado and Romo (2009)]

Let $J \in [2, n]$. The band depth measure for the functions $x_1, \ldots, x_n$ is

$$BD_{n,J}(x) = \sum_{j=2}^{J} BD_n^{(j)}(x). \tag{2.3}$$

For the finite-dimensional version of this measure is important to mention the concept of parallel coordinates, see López-Pintado and Romo (2009), that allows to represent data in $\mathbb{R}^d$. Given a data set $\mathbf{x} \in \mathbb{R}^d$ the observations can be represented as real functions on an index k, $\mathbf{x} = \mathbf{x}(k),\ k \in \{1, \ldots, d\}$. Then the band $B(\mathbf{x}_1, \ldots, \mathbf{x}_j) = \big\{(k, y) : k \in \{1, 2, \ldots, d\}, \min_{i=1,\ldots,j} x_i(k) \leq y \leq \max_{i=1,\ldots,j} x_i(k)\big\}$ is now a d-dimensional interval given by

$$Q(x_1, \ldots, x_j) = \left\{\mathbf{x} \in \mathbb{R}^d : \min_{i=1,\ldots,j} x_i(k) \leq x(k) \leq \max_{i=1,\ldots,j} x_i(k)\right\}. \tag{2.4}$$

In López-Pintado and Romo (2009) the authors use $J = 2$ for several reasons, but mainly because i) $J > 3$ could be computationally expensive and ii) if $J = 2$ even though it is easy to compute, the area of the band can be degenerated with probability one if two curves cross in a given point, then no other curve could be inside the band for that point. The idea behind the band-depth measure is, given a set of curves compute all the combinations of three curves, each combination will define a band. Then count all the curves that are included in each band. The curve that belongs to more bands is the deepest one.

As it was described in Section 2.1, Zuo and Serfling (2000) proposed four properties that a depth function must satisfy. In that sense the finite-dimensional version of the band depth satisfy properties from 2 to 4, the affine invariance property is not satisfied. For a formal proof of these properties and its functional versions see, López-Pintado and Romo (2009), Sections 3 and 4 respectively.

2.2.2 The modified band depth measure

The modified band-depth is a more flexible method to measure the depth of a curve given a functional data set. The flexibility is given because the indicator function in the band-depth measure, that shows when a given function is inside the band, is replaced by the 'proportion of time' that a curve is in the band, through the Lebesgue measure. Formally, for $2 \leq J \leq n$ and for any $x \in X$, let be $A_j(x)$, the interval in I where x is inside the band formed by $x_{i1}, \ldots, x_{ij}$,

$$A_j(x) \equiv A(x; x_{i2}, \ldots, x_{ij}) \equiv \left\{t \in I : \min_{r=i_1,\ldots,i_j} x_r(t) \leq x(t) \leq \max_{r=i_1,\ldots,i_j} x_r(t)\right\}. \tag{2.5}$$

Instead of taking an indicator function that is equal to one when the function x is in the band, a measure of the time that this occurs is proposed by introducing the following ratio,

$$MBD_n^j(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < \cdots < i_j \le n} \lambda_r(A_j(x)), \tag{2.6}$$

where $\lambda_r(A_j(x)) = \dfrac{\lambda(A_j(x))}{\lambda(I)}$ is the proportion of time that the function x is in the band. The total measure of the set I is expressed by the Lebesgue measure λ on I. If the function is always inside the band that ratio takes value one as in the previous case of the band-depth.

Definition 2.6. The modified band-depth [López-Pintado and Romo (2009)].

Let $J \in [2, n]$. The band depth measure for the functions $x_1, \dots, x_n$ is

$$MBD_{n,J}(x) = \sum_{j=2}^{J} MBD_n^j(x). \tag{2.7}$$

For the finite-dimensional case, $MBD_n^j(\mathbf{x})$ is given by

$$MBD_n^j(\mathbf{x}) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < \cdots < i_j \le n} \frac{1}{d} \sum_{k=1}^{d} \mathbf{I}\{min\{x_{i1}(k), \dots, x_{ij}(k)\} \le x(k) \le max\{x_{i1}(k), \dots, x_{ij}(k)\}\}, \tag{2.8}$$

where $\mathbf{x} \in \mathbb{R}^d$ is a multivariate data set. In this case, some of the properties can be proven through the relationship between definition 2.6 and the simplicial depth, see Liu et al. (1990) and López-Pintado and Romo (2009).

For this version in López-Pintado and Romo (2009) the authors consider $J = 2$ because it is computationally fast and also the results are stable with respect to J. The idea behind the modified version of the band depth is for a sample of n curves, consider 'bands' defined for combinations of 2 curves, and account for the proportion of 'x' axis coordinate that a curve c_l is contained in the band (depth index).

Hence, the depth of c_l is defined as the average of the depth index for all the possible bands. The deepest curve is the curve with the maximum depth.

2.2.3 The half-region and modified half-region depth

The *half-region* and *modified half-region* López-Pintado and Romo (2011) depth can be consider as the functional version of the *half-space* depth, and take into account the hypograph and epigraph of a functional datum (curve). In the same setting as for the band depth and modified band depth, (see 2.2.1 and 2.2.2). The hypograph (hyp) and epigraph (epi) of a function (curve) in $x \in C(I)$ is defined as:

$$hyp(x) = \{(t, y) \in I \times \mathbb{R} : y \leq x(t)\}, \tag{2.9}$$

$$epi(x) = \{(t, y) \in I \times \mathbb{R} : y \geq x(t)\} \tag{2.10}$$

Definition 2.7. The half-region depth [López-Pintado and Romo (2011)].

The half-region depth for a function x with respect to a set of functions $x_1, \ldots, x_n$ is

$$HRD(x) = min\left\{\frac{1}{n}\sum_{i=1}^{n} I(G(x_i) \subset hyp(x)), \frac{1}{n}\sum_{i=1}^{n} I(G(x_i) \subset epi(x))\right\} \tag{2.11}$$

Given the previous definition the half-region depth for a function (curve) x_i computes the proportion of curves whose graph belongs to the hypograph of x_i, and the epigraph of x_i, and then takes the minimum value. In the same article the authors presented a less restrictive version of this depth measure, which is based on the 'proportion of time' that the process $X(t) = \{x_1(t), \ldots, x_2(t)\}$ is smaller and greater than de curve x_i. To define that, the authors use the lebesgue measure λ on $\mathbb{R}$, and the respective superior lentgh (SL) and inferior length (IL) constructed as:

$$SL(x) = \frac{1}{\lambda(I)} E[\lambda\{t \in I : x(t) \leq X(t)\}], \tag{2.12}$$

$$IL(x) = \frac{1}{\lambda(I)} E[\lambda\{t \in I : x(t) \geq X(t)\}] \tag{2.13}$$

Definition 2.8. The modified half-region depth [López-Pintado and Romo (2011)] The modified half-region depth for a function x with respect to a set of functions $x_1, \ldots, x_n$ is

$$MHRD(x) = min\left\{\frac{1}{n\lambda(I)}\sum_{i=1}^{n}\lambda\{t \in I : x(t) \leq X(t)\}, \frac{1}{n\lambda(I)}\sum_{i=1}^{n}\lambda\{t \in I : x(t) \leq X(t)\}\right\} \tag{2.14}$$

2.2.4 The random Tukey depth

The random Tukey depth (RTD), is a random approximation of the Tukey depth or halfspace depth. It considers all possible one-dimensional projections of the curves using the halfspace depth. Let $(\mathcal{F}, d) = (\mathbb{H}, \|\cdot\|_{L_2})$, and define $\mathcal{U} = \{u_1, \ldots, u_k\}$ each one sampled independently from a nondegenerate probability measure μ in $\mathbb{H}$, the random Tukey depth is defined as:

Definition 2.9. The random Tukey depth [Cuesta-Albertos and Nieto-Reyes (2008)].

The random Tukey depth for a function x with respect to the probability distribution P is

$$RTD(x, P) = \min_{u \in \mathcal{U}} D_1(\langle u, x\rangle, P_u), \tag{2.15}$$

where for each probability measure Q in a Borel set $\mathbb{R}$, $D_1(t, Q) = min\{Q(-\infty, t], Q[t, -\infty)\}$ and P_u is the marginal of P. The sample version is obtained by substiting P by P_n. This depth function is a random variable in itself, then for the same functional data set can take diferent values, and then order the data in different ways. To be more stable the number of random projections must be increased. For further details see Cuesta-Albertos and Nieto-Reyes (2008).

2.2.5 The h-mode depth

The *h-mode* depth considers the deepest functional datum taking the average of the kernelized distances using the L_2 norm. Let $(\mathcal{F}, d) = (\mathbb{H}, \|\cdot\|_{L_2})$, the h-mode depth is defined as:

Definition 2.10. The h-mode depth [Cuevas et al. (2007)].

The h-mode depth for a function $x \in \mathbb{H}$ with respect to the probability distribution P is

$$h - MD(x, P) = \mathbb{E}K_h(\|x - X\|_{L_2}) \tag{2.16}$$

where $h > 0$, and $K_h(\cdot) = \frac{1}{h}K(\cdot/h)$ as the Gaussian kernel. The sample version with respect to P_n is obtained by taking the sample average:

$$h - MD(x, P_n) = \frac{1}{n}\sum_{i=1}^{n} K_h(\|x - X_i\|_{L_2}) \tag{2.17}$$

$$h - MD(x, P_n) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{\|x - x_i\|_{L_2}}{h}\right) = \sum_{i=1}^{n} K_h(\|x - x_i\|) \tag{2.18}$$

For further details and consistency proofs see Cuevas et al. (2007) and Nagy (2015).

2.2.6 Other functional depths

In the literature there are other several contributions to the concept of functional depth. For instance, the Fraiman-Muniz depth (Integrated depth) proposed by Fraiman and Muniz (2001) measures the conditional quantile on all points. Moreover when the modified band depth is computed with $J = 2$, which is the value used in López-Pintado and Romo (2009), this measure and the integrated depth coincides Nieto-Reyes et al. (2016). The functional spatial depth (FSD) Chakraborty and Chaudhuri (2014) is the extension of the spatial depth from $\mathbb{R}^d$ into infinite-dimensional spaces, and computes the spatial median based on the notion of spatial quantile.

3 K-depth measures for functional data

We start from an available set of sample curves $C = \{\hat{c}_1, \ldots, \hat{c}_m\}$, where $\hat{c}_l \equiv \{(\mathbf{x}_{il}, y_{il}) \in X \times Y\}_{i=1}^n$, where X is a compact subset of $\mathbb{R}^n$ and, in most cases, $Y = \mathbb{R}$. We can assume that the $\mathbf{x}_i's$ are common for all the curves, and that for each $\hat{c}_l$, exists a continuous function $c_l : X \longrightarrow Y$ such that $E[y_l|\mathbf{x}] = c_l(\mathbf{x})$ (with respect to a given probability measure).

These functions are the functional data curves and can be considered as points in some functional space. We will project these points onto some finite-dimensional function subspace, in our case, a Reproducing Kernel Hilbert Space (RKHS), H_K, generated by a Mercer kernel K. Consider the integral operator T_K defined by $T_K(f) = \int_X K(\cdot, s)f(s)ds$. T_K has a countable sequence of eigenvalues $\{\lambda_j\}$ and (orthonormal) eigenfunctions $\{\phi_j\}$ and K can be expressed as $K(x, y) = \sum_j \lambda_j \phi_j(x)\phi_j(y)$ where the convergence is absolute and uniform (Mercer Theorem).

Given a function f in a function space containing H_K, it will be projected onto H_K using the operator T_K. By the Spectral Theorem, the projection $f^* = T_K(f) \in H_K$ takes the form $f^* = T_K(f) = \sum_j \lambda_j \langle f, \phi_j\rangle \phi_j$. To determine the $\langle f, \phi_j\rangle$ coefficients, we solve the Support Vector Machine (SVM) regularization problem:

$$\arg \min_{c \in H_K} \frac{1}{n} \sum_{i=1}^{n} L(y_i, c(\mathbf{x}_i)) + \gamma \|c\|_K^2 \ ,$$

where $\gamma > 0$, $\|c\|_K$ is the norm of the function c in H_K, $y_i = \hat{c}_i$ and $L(y_i, c(\mathbf{x}_i)) = (|c(\mathbf{x}_i) - y_i| - \varepsilon)_+$, $\varepsilon \geq 0$, see Moguerza and Muñoz (2006).

The Representer Theorem (see Cucker and Smale (2002)) states that the solution to this optimization problem is given by $c_l^*(\mathbf{x}) = \sum_{i=1}^n \alpha_{il} K(\mathbf{x}_i, \mathbf{x})$, $\forall \mathbf{x} \in X$, where $\alpha_{il} \in \mathbb{R}$ are the Lagrange multipliers associated to the support vectors.

Let c_l be a curve, whose sample version is $\hat{c}_l \equiv \{(\mathbf{x}_{il}, y_{il}) \in X \times Y\}_{i=1}^n$. Consider the functional representation for c_l given by $\lambda_l^* = (\lambda_{1l}^*, \ldots, \lambda_{dl}^*)$, where

$$\lambda_{jl}^* = \sum_{i=1}^{n} \hat{\lambda}_j \alpha_{il} \hat{\phi}_{ji} \ , \tag{3.1}$$

α_{il} are given by the solution of the SVM, $\hat{\lambda}_j$ is the eigenvalue corresponding to the eigenvector $\hat{\phi}_j$ of the matrix $K_S = (K(\mathbf{x}_i, \mathbf{x}_j))_{i,j}$, and $d = \min(n, rank(K_S))$. Now $f^* = T_K(f) = \sum_j \lambda_j \langle f, \phi_j \rangle \phi_j = \sum_j \lambda_j^* \phi_j \simeq \sum_j \left(\hat{\lambda}_j \sum_{i=1}^n \alpha_i \hat{\phi}_{ji} \right) \phi_j$ and $\sum_{i=1}^n \alpha_i \hat{\phi}_{ji} \simeq \langle f, \phi_j \rangle$, (see Muñoz and González (2010) for further details).

Once we are able to represent each curve as a point in $\mathbb{R}^d$ we want to exploit this functional representation of a set of functional data by defining depth measures associated to this representation. A very natural way is to consider the multivariate median computed coordinatewise and consider a distance function from each point to the median.

Definition 3.1. K-deepest point. Given a set of sample functional data C and the corresponding H_K-representations $\lambda_l^* \equiv (\lambda_{1l}^*, \ldots, \lambda_{dl}^*)$, we define the K-depeest functional data point as the multivariate median of the d-dimensional functional data points, computed as the vector of the coordinatewise medians in $\mathbb{R}^d$: $\mathbf{P}^* = (p_1^*, \ldots, p_d^*)$, where $p_i^* = median\{\lambda_{il}^*\}$.

Following the Mahalanobis depth measure presented in def. 2.2 which is based on the so-called distance, we present a kernelized version of the Mahalanobis depth adapted to the representation we have defined.

Definition 3.2. Kernel Mahalanobis Depth. Given the d-dimensional K-deepest point $\mathbf{P}^*$, the Kernel Mahalanobis Depth (KMD) from a functional data point $\hat{c}_l$ to $\mathbf{P}^*$ is defined as theinverse of the Mahalanobis distance between the H_K-representation of $\hat{c}_l$ and $\mathbf{P}^*$:

$$KMD(\hat{c}_l, \mathbf{P}^*) = \frac{1}{[(\lambda_l^* - \mathbf{P}^*)^T \mathbf{\Sigma}_{\lambda^*}^{-1} (\lambda_l^* - \mathbf{P}^*)]^{-1/2}}, \tag{3.2}$$

where $\mathbf{\Sigma}_{\lambda^*}^{-1}$ is the inverse of covariance matrix of the functional data set (computed from its H_K-representation).

As it was mentioned in Section 2.1 the Mahalanobis depth satisfies the four properties mentioned only in the case when the probability distribution F is symmetric. In this sense in Martos et al. (2014) the authors shown that the Mahalanobis distance does not work well under two situations: first, when the underlying distribution of the data is not Gaussian, the key property of the Mahalanobis distance is not preserved:

Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$ with a probability distribution $\mathbb{P}$ and $\{\mathbf{x_1}, \cdots, \mathbf{x_n}\}$ be a random sample of $\mathbf{X}$, then if $f_{\mathbb{P}}(\mathbf{x}_i) = f_{\mathbb{P}}(\mathbf{x}_j) = c$, then $\mathbf{x}_i$ and $\mathbf{x}_j$ are equally distant from the center of the distribution, for all $i \neq j$, where $f_{\mathbb{P}}$ is the density function of $\mathbf{X}$. In other words $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same level curve of $f_{\mathbb{P}}(\cdot)$. Therefore as the Mahalanobis depth is the inverse of the Mahalanobis distance, it also fails in this case. Second, to derive a sample version of Mahalanobis distance an estimation of the covariance matrix is needed, which can lead to some problems when dimensionality increases or there are outliers, see (Zhang et al., 2012, Martos et al., 2014).

To solve that issue Martos et al. (2014) proposed a generalization of the Mahalanobis distance via density kernels. This involve defining a family of kernels based on the underlying density function that is presented in the data. Previously to the definition of the density Kernel and the generalization of the Mahalanobis distance is important to introduce the concept asymptotic of f-monotonicity.

Definition 3.3. (asymptotic f-monotonicity). Consider a random sample $S_n = \{\mathbf{x}_i\}_{i=1}^n$ drawn from $\mathbb{P}$. A function $g(\mathbf{x}, S_n)$ is asymptotically f-monotone if:

$$f_{\mathbb{P}}(\mathbf{x}) \geq f_{\mathbb{P}}(\mathbf{y}) \Rightarrow \lim_{n\to\infty} P(g(\mathbf{x}, S_n) \geq g(\mathbf{y}, S_n)) = 1.$$

Definition 3.4. Density Kernel. Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$ with a probability distribution $\mathbb{P}$ and let $g(x, \mathbb{P})$ be a positive f-monotone function (see 3.3). Define $\phi_{\mathbb{P}} : \mathbf{X} \to \mathbb{R}^+$ as $\phi_{\mathbb{P}}(\mathbf{x}) = g(\mathbf{x}, \mathbb{P})$. The density kernel is defined as:

$$K_{\mathbb{P}}(\mathbf{x}, \mathbf{y}) = \phi_{\mathbb{P}}(\mathbf{x})\phi_{\mathbb{P}}(\mathbf{y}) \tag{3.3}$$

It can be proven that $K_{\mathbb{P}}$ is a Mercer Kernel, hence the generalized Mahalanobis (GM) distance can be defined as follows.

Definition 3.5. Generalized Mahalanobis distance associated to $K_{\mathbb{P}}$. Let $\mathbf{X}$ be a random vector in $\mathbb{R}^d$ with a probability distribution $\mathbb{P}$ and let $K_{\mathbb{P}}(\mathbf{x}, \mathbf{y})$ be a density kernel. The GM distance associated to the density kernel $K_{\mathbb{P}}$, from a point $\mathbf{x}$ to the center of the distribution $\mathbb{P}$, is:

$$d^2_{GM_{K_{\mathbb{P}}}}(\mathbf{x}, \mathbf{m_0}) = -log K_{\mathbb{P}}(\mathbf{x}, \mathbf{m_0}), \tag{3.4}$$

where $\mathbf{m_0}$ is the point of maximum density of the distribution, $\mathbf{m_0} = arg \max_{\mathbf{x}} f_{\mathbb{P}}(\mathbf{x})$. For further details and properties see Martos et al. (2013).

Next we propose a generalized kernel depth base on this generalized Mahalanobis distance.

Definition 3.6. Generalized Kernel Depth. The generalized kernel (GK) depth from a functional data point $\hat{c}_l$ with respect to a functional data set C is defined as the exponential of the negative of the generalized Mahalanobis distance between the $\{\lambda_l^*\}$ (the H_K-representation of $\hat{c}_l$, given in def. 3.1) and $\mathbf{m_0}$:

$$\begin{aligned} GKD(\hat{c}_l, C) &= exp\{-d^2_{GM_{K_{\mathbb{P}}}}(\lambda_l^*, \mathbf{m_0})\} \\ &= exp\{log K_{\mathbb{P}}(\lambda_l^*, \mathbf{m_0})\} \\ &= K_{\mathbb{P}}(\lambda_l^*, \mathbf{m_0}), \\ &= \phi_{\mathbb{P}}(\lambda_l^*)\phi_{\mathbb{P}}(\mathbf{m_0}), \end{aligned} \tag{3.5}$$

where $\lambda_l^* \equiv (\lambda_{1l}^*, \ldots, \lambda_{dl}^*)$ is the H_K-representation of $\hat{c}_l$ and $\mathbf{m_0}$ is the point of maximum density of the distribution of λ_l^*, $\mathbf{m_0} = arg \max_{\boldsymbol{\lambda}_\mathbf{l}^*} f_{\mathbb{P}}(\lambda_l^*)$.

In next Section we prove the fullfilment of properties 1-4 in Section 2.1. As we show in the experimental Section (see table 2) the h-mode depth and the GK depth perform similarly for some data sets. This motivates our following proposition.

Proposition 3.1. Generalizing the h-mode depth.

The h-mode depth is a particular case of the generalized kernel depth.

Proof.

Using the previous notation, let $\mathbf{x}$ be a sample curve and $\lambda^*_{\mathbf{x}}$ its H_K representation. The sample version of the h-MD is given by $h - MD(\mathbf{x}, C) = \sum_{i=1}^{m} K_h(\|\mathbf{x} - \hat{c}_i\|)$, with $C = \{\hat{c}_1, \ldots, \hat{c}_m\}$.

Let $\phi_{\mathbb{P}}(\mathbf{x}) = \sum_{i=1}^{m} K_h(\|\mathbf{x} - \hat{c}_i\|)$, then by definition of the generalized kernel depth

$$
\begin{aligned}
GKD(\mathbf{x}, C) &= \phi_{\mathbb{P}}(\lambda^*_{\mathbf{x}})\phi_{\mathbb{P}}(\mathbf{m_0}), \\
&= \left(\sum_{i=1}^{m} K_h(\|\lambda^*_{\mathbf{x}} - \hat{c}_i\|)\right)\left(\sum_{i=1}^{m} K_h(\|\mathbf{m_0} - \hat{c}_i\|)\right).
\end{aligned}
$$

Using the normalized version of the h-mode depth $\frac{h-MD(\mathbf{x},C)-min(h-MD(\mathbf{x},C))}{max(h-MD(\mathbf{x},C))-min(h-MD(\mathbf{x},C))}$, as $\mathbf{m_0} = arg \max_{\lambda^*_{\mathbf{x}}} f_{\mathbb{P}}(\lambda^*_x)$ then $\mathbf{x^{m_0}}$ is the curve associated to $\mathbf{m_0}$, hence $max(h - MD(\mathbf{x}, C)) = h - MD(\mathbf{x^{m_0}}, C)$ and thus $\frac{h-MD(\mathbf{x^{m_0}},C)-min(h-MD(\mathbf{x},C))}{max(h-MD(\mathbf{x},C))-min(h-MD(\mathbf{x},C))} = \sum_{i=1}^{m} K_h(\|\mathbf{m_0} - \hat{c}_i\|) = 1$. Therefore,

$$
GKD(\mathbf{x}, C) = \sum_{i=1}^{m} K_h(\lambda^*_{\mathbf{x}} - \hat{c}_i\|)
$$

As $\lambda^*_{\mathbf{x}}$ its H_K representation of the sample curve $\mathbf{x}$, then the h-mode depth is a particular case of the generalized kernel depth. It remains to proof that the h-mode depth is a density kernel with respect to the density of the projected curves.

As $K_h(\cdot)$ is a kernel density estimator, it can be proven that converges in probability to $f_{\mathbb{P}}(C)$, where $f_{\mathbb{P}}(C)$ corresponds to the density function of the curves projected onto RKHS defined previously as $f_{\mathbb{P}}(\lambda^*_{\mathbf{x}})$. Therefore $\phi_{\mathbb{P}}(\mathbf{x}) = \sum_{i=1}^{m} K_h(\|\mathbf{x} - \hat{c}_i\|)$ is a positive asymprotic f-monotone function, what concludes the proof. ■

3.1 Properties of the Generalized Kernel depth

In this Section we proof that the GK depth satisfies properties 1-4 described in Section 2.1.

Proposition 3.2. The order induced by the GK depth in the H_K representation of the curves is affine invariant.

Consider the rank mapping $I : C \to \mathbb{N}$ that ranks the $\hat{c}_l$ curves according to their GK depth values. Then

$I(\hat{c}_l) = I(f(\hat{c}_l))$, for $f(\mathbf{x}) = a + b\mathbf{x}$ (affine transformation). Notice that we identify $\hat{c}_l$ with the H_K representation (λ_l^*). Thus $f(\hat{c}_l)$ refers to $f(\lambda_l^*)$.

Proof. Let $\hat{c}_l = \{(\mathbf{x}_{il}, y_{il}) \in X \times Y\}_{i=1}^n$ then $f(\hat{c}_l) = (\mathbf{x}_{il}, f(y_{il}))$. Consider the loss function,

$$
\begin{aligned}
L(y_i, c(\mathbf{x}_i)) &= (|c(\mathbf{x}_i) - y_i| - \varepsilon)_+, \ \varepsilon \geq 0, \text{ then} \\
L(f(y_i), f(c(\mathbf{x}_i))) &= (|f(c(\mathbf{x}_i)) - f(y_i)| - \varepsilon)_+, \\
&= (|a + bc(\mathbf{x}_i) - a - by_i| - \varepsilon)_+, \\
&= (|b(c(\mathbf{x}_i) - y_i)| - \varepsilon)_+, \\
&= |b|(|(c(\mathbf{x}_i) - y_i)| - \varepsilon)_+, \\
&= |b|L(y_i, c(\mathbf{x}_i)).
\end{aligned}
$$

Therefore $f(\hat{c}_l) = |b|\hat{c}_l$. As $\hat{c}_l$ can be approximated by $\sum_j \lambda_j^* \phi_j$, then $f(\hat{c}_l) = f(\sum_j \lambda_j^* \phi_j) = \sum_j |b|\lambda_j^* \phi_j = \sum_j \gamma_j^* \phi_j$, with $\gamma_j^* = |b|\lambda_j^*$. As $|b| \geq 0 \Rightarrow I(\hat{c}_l) = I(f(\hat{c}_l))$ ■

Proposition 3.3. The GK depth presents maximality at center.
$GKD(\hat{c}_l^{m_0}, C) = \sup_{\hat{c}_l} GKD(\hat{c}_l, C)$ for any $\hat{c}_l^{m_0}$ that is the center of C.

Proof. Consider $\mathbf{m_0} = arg \max_{\lambda_l^*} f_{\mathbb{P}}(\lambda_l^*)$, and let $\hat{c}_l^{m_0}$ be the curve associated to $\mathbf{m_0} \Rightarrow D_{GK}(\hat{c}_l^{m_0}, C) = K_{\mathbb{P}}(\mathbf{m_0}, \mathbf{m_0}) = 1$. Therefore by def. 3.6 $GKD(\hat{c}_l, C)$ takes its maximum at $GKD(\hat{c}_l^{m_0}, C)$ ■

Proposition 3.4. The GK depth presents asymptotic monotonicity relative to the depeest point.
For any set of curves C where $\hat{c}_l^{m_0}$ is the deepest functional datum, $GKD(\hat{c}_l, C) \leq GKD(\hat{c}_l^{m_0} + \alpha(\|\hat{c}_l - \hat{c}_l^{m_0}\|), C) \leq GKD(\hat{c}_l^{m_0}, C)$, for $\alpha \in (0, 1)$.

Proof. By definition 3.4 and 3.6, $GKD(\hat{c}_l, C) = K_{\mathbb{P}}(\lambda_l^*, \mathbf{m_0}) = \phi_{\mathbb{P}}(\lambda_l^*)\phi_{\mathbb{P}}(\mathbf{m_0})$ is a density kernel. Hence it satisfies definition 3.3 that $GKD(\hat{c}_l, C)$ is an asymptotically f-monotone function. Therefore preserves monotonicity with respect to $GKD(\hat{c}_l^{m_0}, C)$ for any $\mathbf{m_0} = arg \max_{\lambda_l^*} f_{\mathbb{P}}(\lambda_l^*)$ ■

Proposition 3.5. The GK depth vanishes at infinity.

$D_{GK}(\hat{c}_l, C) \to 0$ *if* $\|\hat{c}_l\| \to \infty$.

Proof. If $\|\hat{c}_l\| \to \infty$ then $d^2_{GM_{K_{\mathbb{P}}}}(\lambda_l^*, \mathbf{m_0}) \to \infty \Rightarrow D_{GK}(\hat{c}_l, C) = exp\{-d^2_{GM_{K_{\mathbb{P}}}}(\lambda_l^*, \mathbf{m_0})\} \to 0$ ■

In proposition 3.2-3.5 we proved that the generalized kernel depth satisfies the all the necessary properties proposed by Zuo and Serfling (2000) to be a depth function. A last crucial question hast to be answered. As our kernel depths depend on the choice of a particular basis of approximating functions, will the rank induced on the curves set change when we chose a different basis? The answer is no:

Theorem 3.1. The order induced by the GK depth is asymptotically invariant to the RKHS basis choice.

$I(\hat{c}_l)$ is asymptotically invariant for all orthonormal basis as defined in Section 3.

Proof. *Let $\hat{c}_l \equiv \{(\boldsymbol{x}_{il}, y_{il}) \in X \times Y\}_{i=1}^n$ be a curve of c. Consider two othonormal basis $B_1 = \{\phi_i\}$ and $B_2 = \{\psi_i\}$ (defined in Section 3).*

Then $\hat{c}_l$ can be approximated by both $\hat{c}_{B_1} \equiv \sum_{i=1}^{l} \lambda_i^* \phi_i$ and $\hat{c}_{B_2} = \sum_{j=1}^{h} \gamma_j^* \psi_j$, for $l, h < \infty$ (given that $n < \infty$) and where $\hat{c}_{B_1} \approx \hat{c}_{B_2}$ ($\|\hat{c}_{B_1} - \hat{c}_{B_2}\| < \epsilon$, for some acceptable ϵ). Consider the functional subspaces $Span(\phi_i)$, $Span(\psi_j)$. Each function $\{\phi_i\}$, $\{\psi_j\}$ constitutes a basis on a functional (Hilbert) subspace included in L^2, namely the closure of $Span(\phi_i)$, $Span(\psi_j)$. Therefore $\{\phi_i\}$ can be expressed in terms of $\{\psi_j\}$ or in the other way (from $\{\psi_j\}$ to $\{\phi_i\}$).

Then $\hat{\phi}_i \approx \sum_{j=1}^{h^*} \alpha_{ij} \psi_j$. Then as we know that $\hat{c}_{B_1} \equiv \sum_{i=1}^{l} \lambda_i^* \phi_i$, therefore $\hat{c}_{B_1} \equiv \sum_{i=1}^{l} \lambda_i^* \sum_{j=1}^{h^*} \alpha_{ij} \psi_j = \sum_{i=1}^{l} \sum_{j=1}^{h^*} \omega_{ij}^* \hat{\psi}_j$, where $\omega_{ij}^* = \lambda_i^* \alpha_{ij}$

Therefore we can go from $\hat{c}_{B_1}$ to $\hat{c}_{B_2}$ by a change of basis, a particular case of affine transformation. Therefore by proposition 3.2 $I(\phi_{\mathbb{P}}(\lambda_l^*) \phi_{\mathbb{P}}(\mathbf{m_0})) = I(\phi_{\mathbb{P}}(\omega_{ij}^*) \phi_{\mathbb{P}}(\mathbf{m_0}))$, ■

The next example ilustrates Theorem 3.1. In the following figures are shown the first two RKHS projections of the data using the RBF kernel (left panel), the first two RKHS projections of the data using the

Spline kernel (middle panel), and the order induced by the generalized kernel depth using two different basis functions, in the 'y' axis are considered the eigenfunctions of the RBF kernel in spline basis and in the 'x' axis is considered the eigenfunctions of the spline kernel (right panel). The verification excercise was made for the Australian Mortality rates data, that contains age-specific mortality rates for Australian males for 1901-2003 in logarithmic scale, which is publicly available in the R package 'fds'; for the Berkeley growth study data (boys and girls) and for the Vertical Density Profile data (see Section 4 for details). As can be observed the order in all the cases remains unchanged, in other words the change of basis does not alter the order induced as was proven in Theorem 3.1.

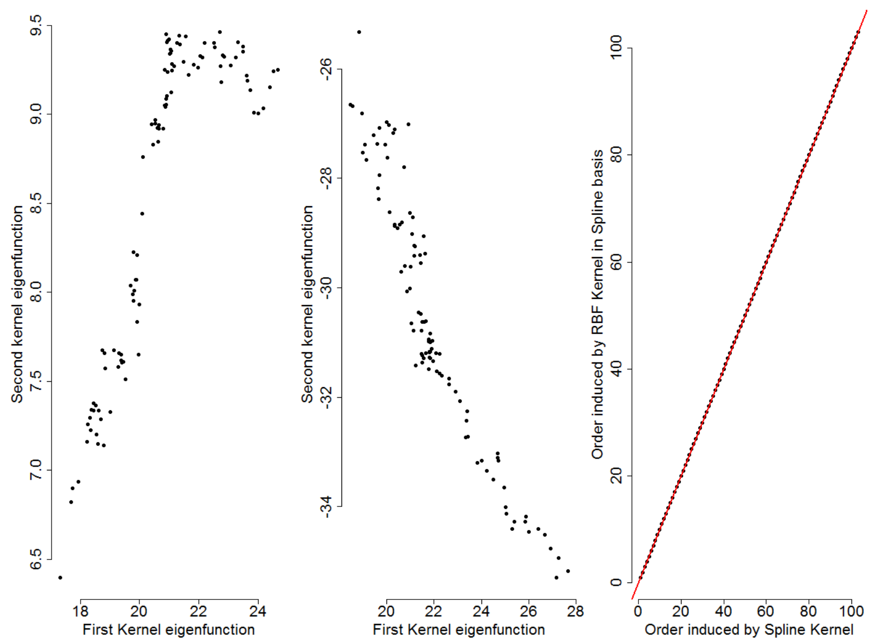

Figure 1: **Australian mortality rates data (103 years)**. First two RKHS projections for RBF ($\sigma = 0.1$) and Spline kernels respectively and the order induced by the GKD applying change the of basis procedure.

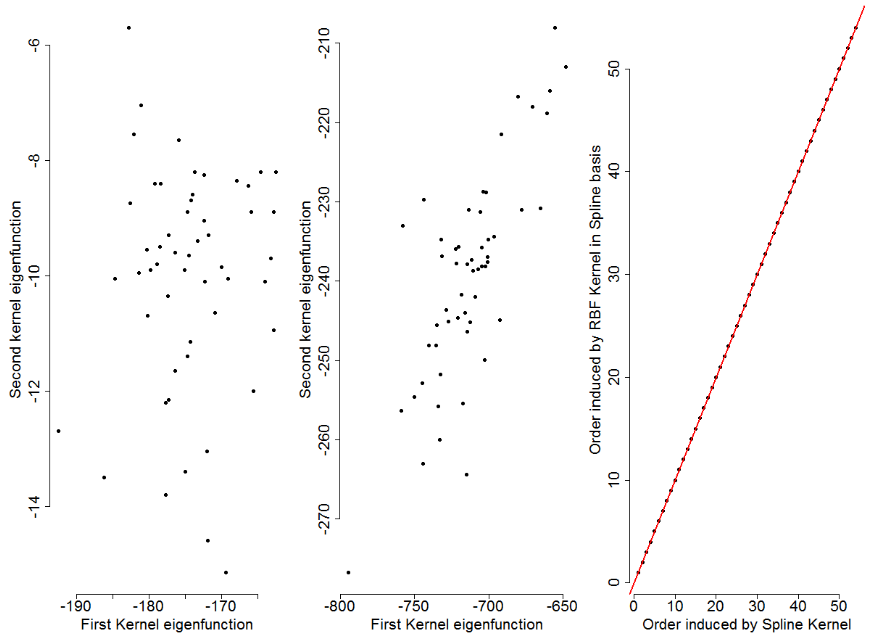

Figure 2: **Berkeley growth study data: (54 girls)**. First two RKHS projections for RBF ($\sigma = 0.015$) and Spline kernels respectively and the order induced by the GKD applying change the of basis procedure.

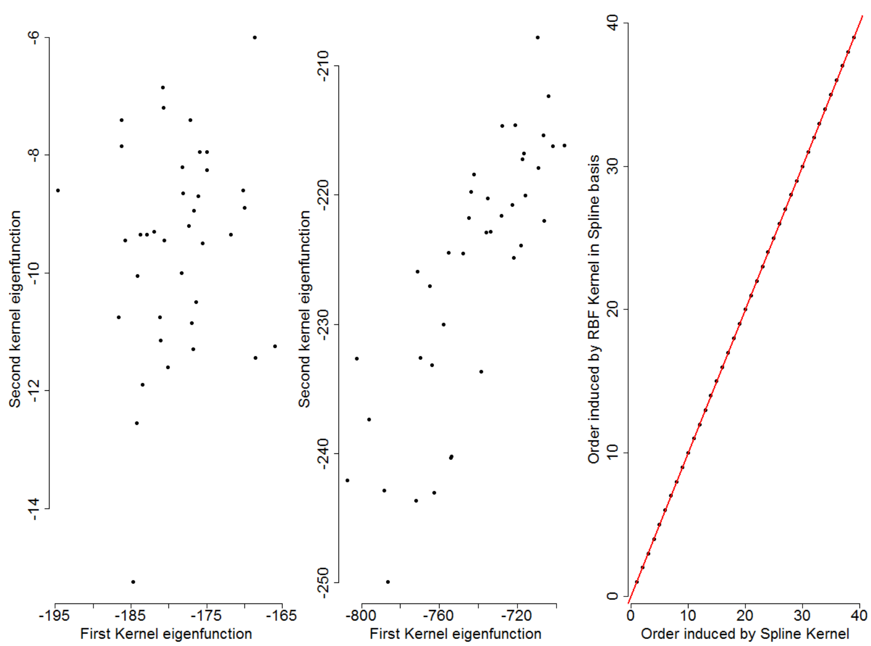

Figure 3: **Berkeley growth study data (39 boys)**. First two RKHS projections for RBF ($\sigma = 0.015$) and Spline kernels respectively and the order induced by the GKD applying change the of basis procedure.

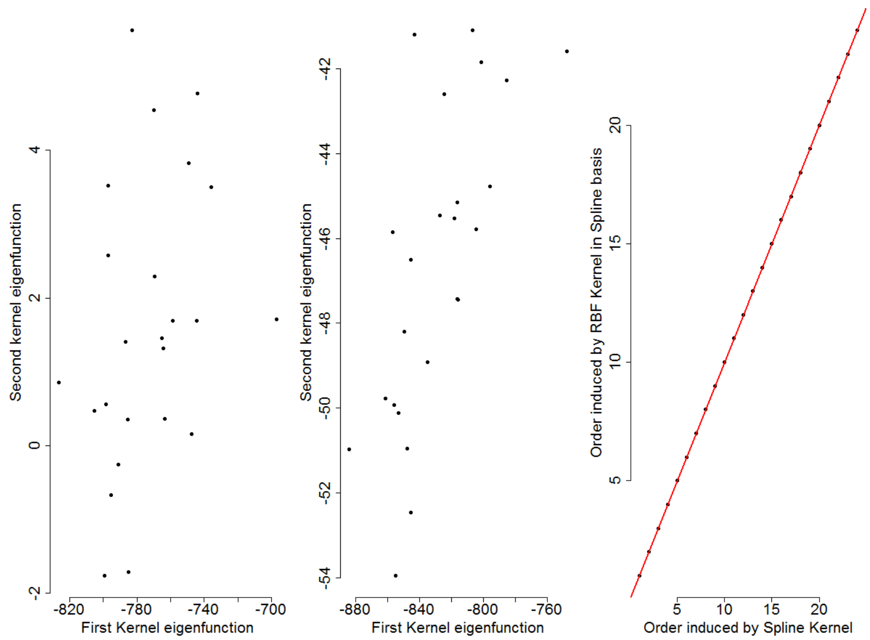

Figure 4: **Vertical Density Profile data (24 profiles)**. First two RKHS projections for RBF ($\sigma = 0.002$) and Spline kernels respectively and the order induced by the GKD applying change the of basis procedure.

3.2 Polynomial Kernels

As it was stated in the previously the H_K representation of the functional datum $\hat{c}_l$ is obtained by the integral operator T_K defined by $T_K(f) = \int_X K(\cdot, s)f(s)ds$. In this sense each curve needs a kernel function to be projected onto a RKHS. In this paper we consider different and typical kernel functions, such as the Gaussian kernel, the polynomial kernel and the spline kernel. Nevertheless given that $K(x, y) = \sum_j \lambda_j \phi_j(x)\phi_j(y)$, several kernel functions can be constructed by taking different functions ϕ_j such that this functions constitute an orthonormal basis.

Definition 3.7. (Orthogonal Polynomials) Let $\mathcal{A} = \{p_n\}_{n\in\mathbb{N}}$ be a sussecion of functions $\in L^2(\Omega, \rho)$ then $\mathcal{A}$ is a sussecion of orthogonal polynomials if:

$$\langle p_i, p_j \rangle_{L^2(\Omega,\rho)} = \int_\Omega p_i(x)p_j(x)\rho(x)d\mu = \delta_{ij}, \tag{3.6}$$

where μ is a probability measure and $\rho(x)$ is a weight function. Consider $P_n(x)$ as the Legendre polynomial of degree n and the weight function $\rho(x) = 1$, then the Legendre kernel is defined as:

Definition 3.8. Legendre Polynomial Kernel

$$k(x,y) = \langle P_n, P_m \rangle_{L^2([-1,1],\rho)} = \int_{[-1,1]} P_n(x) P_m(y) dx = \frac{(2n+1)}{2n(n+1)} \delta_{nm}. \tag{3.7}$$

This kernel is continuous in the interval $(-1, 1)$ and the term $\frac{(2n+1)}{2n(n+1)}$ is needed to ensure convergence. Consider $Q_n(x)$ as the Laguerre polynomial of degree n and the weight fucntion $\rho(x) = exp\{-x\}$ the Laguerre kernel is definded as:

Definition 3.9. Laguerre Polynomial Kernel

$$k(x,y) = \langle Q_n, Q_m \rangle_{L^2([0,\infty),\rho)} = \int_{[0,\infty)} Q_n(x) Q_m(y) e^{-x} dx = (n!)^2 \delta_{nm}, \tag{3.8}$$

where the kernel function is continuos in the interval $[0, \infty]$. Consider T_n as the Hermite Polynomial of degree n and the weight function $\rho(x) = exp\{\frac{-x^2}{2}\}$, the Hermite kernel is defined as:

Definition 3.10. Hermite Polynomial Kernel

$$k(x,y) = \langle T_n, T_m \rangle_{L^2(\mathbb{R},\rho)} = \int_{\mathbb{R}} T_n(x) T_m(y) e^{\frac{-x^2}{2}} dx = \sqrt{2\pi} n! \delta_{nm}, \tag{3.9}$$

where the kernel function is continuos in $\mathbb{R}$. For further details see Fasshauer and McCourt (2015).

4 Experimental Work

On of the objetives of this paper is to test wether the performance of statistical functional measures (such as depth) is preserved or enhanced when we work with nontrivial functional representations of functional data (as opposed to the measures defined on the plain representation of data). Thus we proposed two kernel depth measures that use the FDA coordinates instead of the plain curve representations. Because depth induces a center-outward ordering of multivariate data sets/curves, a natural problem to test the utility of different depth definitions is outlier detection.

In this Section we present the results of the proposed kernel depths on the task of functional outlier detection, for different sets of simulated curves and real functional data sets. In the multivariate framework an outlier is an observation that lies outside the overall pattern of a distribution, so it depends on the distance from the center of the distribution. In the functional framework can also appear shape outliers, that is, curves which are not far away from the bulk of data, but they present a different shape Moguerza et al. (2007).

To perform the empirical work we choose three different and typical kernel functions, namely: i) Gaussian kernel $K_G(x_i, x_j) = e^{-\sigma\|x_i - x_j\|^2}$; ii) polynomial kernel $K_P(x_i, x_j) = (a\langle x_i, x_j\rangle + b)^d$; iii) spline kernel $K_S(x_i, x_j) = \prod_{d=1}^{D} 1 + x_i x_j + x_i x_j min(x_i, x_j) - \frac{x_i+x_j}{2} min(x_i, x_j)^2 + \frac{x_i+x_j}{3} min(x_i, x_j)^3$. Also we have constructed different kernel functions using the orthonormal Legendre, Laguerre and Hermite porlynomials, see Section 3.2. All the parameters, including the penalization coefficient γ of the SVM regularization problem (to obtain the H_K representations) were defined through cross-validation. We test our depth measures against several well known depth functions described in Section 2 namely: The band Depth (BD), the modified band depth (MBD), the half-region depth (HRD), the modified half-region depth (MHRD), the random Tukey depth (RTD) and the h-mode depth (HMD).

4.1 Artificial data set I.

We simulate 100 curves, 95 drawn from the same population given by the distribution of the coefficients a_i plus 5 curves with a different parametrization in the role of outlying curves. The shape of the two types of curves are different as can be appreciated in Figure 5:

$f_i(x_t) = a_i + 0.05t + sin(\pi x_t^2), \quad i = 1, \dots 95,$

$f_i^o(x_t) = b_i + 0.05t + cos(20\pi x_t), \quad i = 96, \dots 100, \quad \text{(outlying curves)}$

where $x_t = \dfrac{t}{500} \in [0, 1]$, $t = 1, \dots, 500$, $a_i \sim N(\mu_a = 5, \sigma_a = 4)$, $b_i \sim N(\mu_b = 5, \sigma_b = 3)$.

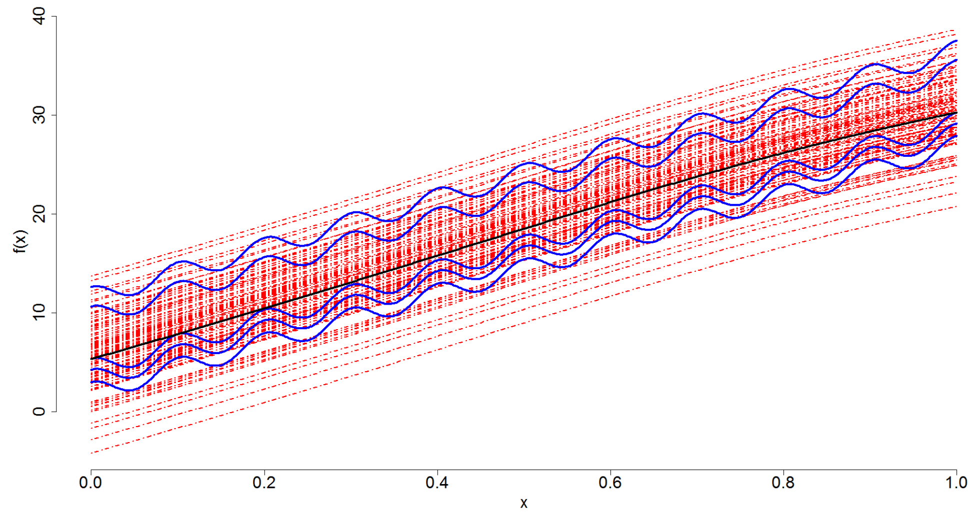

Figure 5: Artificial data set I. The five waved curves are the outliers, and the black line is the deepest curve.

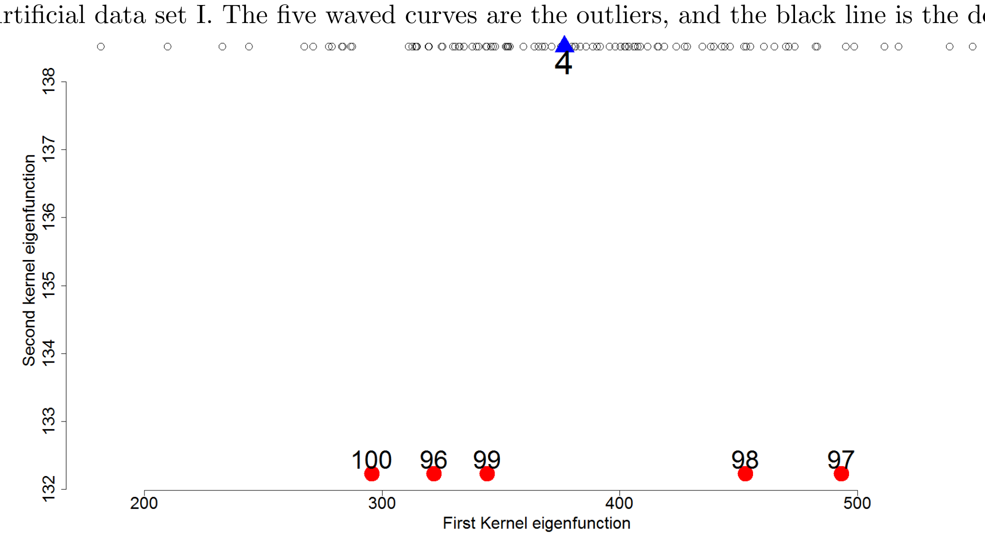

Figure 6: Artificial data set I. RKHS projections. Main population (up-black dots), the outlying curves (down-red dots) and the deepest curve (up-blue triangle).

The results are summarized in table 1. Bothe kernel depths proposed are able to detect exactly the five outlying curves. The competitors techniques fail to capture any of the true outliers, except for the BD and HRD that identify one outlying curve. Moreover in Figure 6 are illustrated the two first projections of the curves onto a functional space. Throughout this representation the data can be perfectly discriminated (down-red points).

Table 1: **Artificial data set I.** Number of outliers, false-Positive and false-Negative identifications ($\gamma = 10^{-7}$).

Measure	True different outliers	False positive	False negative
$GKD_{G_{(\sigma=500)}}$	**5**	**0**	**0**
$GKD_{P_{(a=1,\ b=1,\ d=10)}}$	**5**	**0**	**0**
GKD_s	**5**	**0**	**0**
$GKD_{Leg_{(n=2)}}$	**5**	**0**	**0**
$GKD_{Lag_{(n=2)}}$	**5**	**0**	**0**
$GKD_{Her_{(n=2)}}$	**5**	**0**	**0**
$KMD_{G_{(\sigma=500)}}$	**5**	**0**	**0**
$KMD_{P_{(a=1,\ b=1,\ d=10)}}$	**5**	**0**	**0**
KMD_S	**5**	**0**	**0**
$KMD_{Leg_{(n=2)}}$	**5**	**0**	**0**
$KMD_{Lag_{(n=2)}}$	**5**	**0**	**0**
$KMD_{Her_{(n=2)}}$	**5**	**0**	**0**
BD	1	4	4
MBD	0	5	5
HRD	1	4	4
MHRD	0	5	5
RTD	0	5	5
HMD	0	5	5

4.2 Artificial data set II.

We simulate 108 curves. 100 generated under the process $f_i(t) = sin(4\pi t) + \epsilon(t)$ for $i = 1, \dots 100$, where $t \in [0, 1]$, $\epsilon(t)$ is a Gaussian process with zero mean and covariance function $\gamma(s, t) = 0.2 \times exp\{-0.8|s-t|\}$. Also 8 curves were generated with a different parametrization in the role of outlying curves:

Shape outliers: $f_i^{s_1}(t) = 0.05 \times \omega(t)$; $f_i^{s_2}(t) = 0.25 \times \omega(t)$; $f_i^{s_3}(t) = -0.25 + cos(4\pi t) + 0.05 \times \omega(t)$;$f_i^{s_4}(t) = 0.25 + cos(4\pi t) + 0.05 \times \omega(t)$.

Magnitude outliers:$f_i^{m_1}(t) = 1 + sin(4\pi t) + \omega(t)$; $f_i^{m_2}(t) = 2 + sin(4\pi t) + \omega(t)$; $f_i^{m_3}(t) = -1 + sin(4\pi t) + \omega(t)$; $f_i^{m_4}(t) = -2 + sin(4\pi t) + \omega(t)$;

where ω is an uncorrelated Gaussian process with zero mean and variance = 1. The results are summarized in table 2. In this case both kernel depths are able to detect exactly the eight outlying curves.

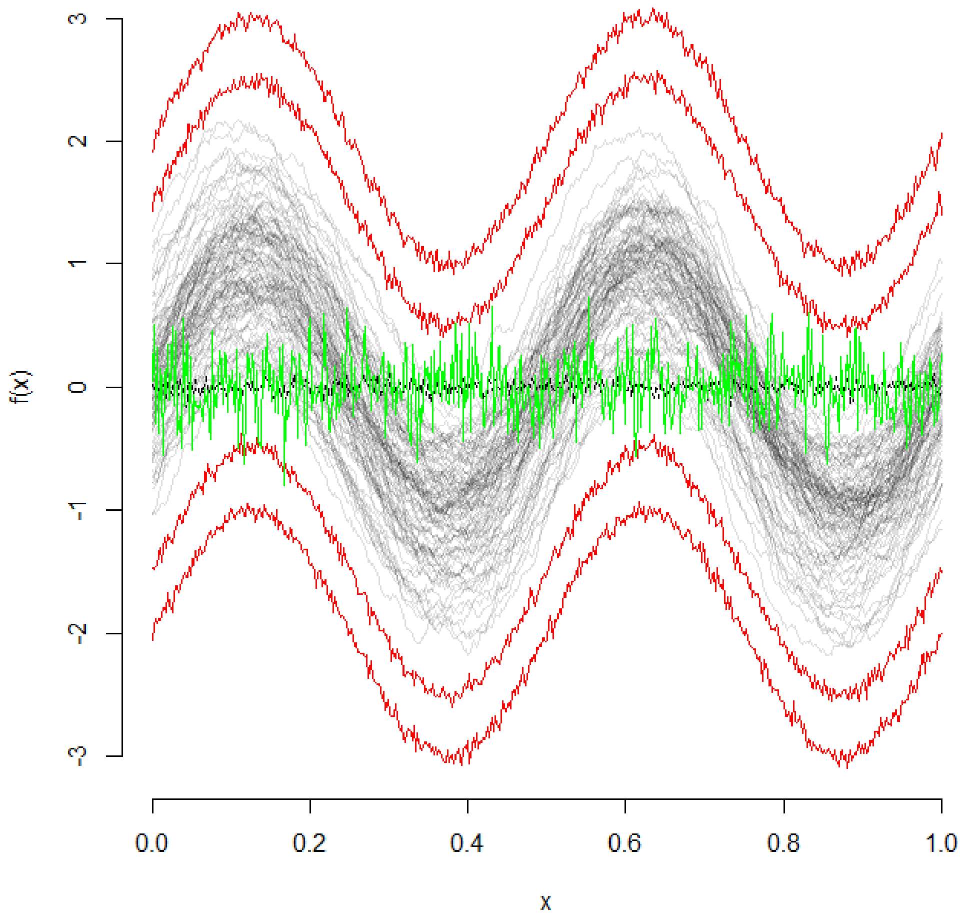

Figure 7: Artificial data set II. The 100 waved curves in grey are the core data. The waved curves in blue and red are the outliers.

Table 2: **Artificial data set II.** Number of outliers correctly identified ($\gamma = 10^{-5}$).

Measure	True different shape outliers	True different magnitude outliers
$GKD_{G_{(\sigma=0.001)}}$	**4**	**4**
$GKD_{P_{(a=2,\ b=2,\ d=10)}}$	**4**	**4**
GKD_s	**4**	**4**
$GKD_{Leg_{(n=8)}}$	**4**	**4**
$GKD_{Lag_{(n=8)}}$	**4**	**4**
$GKD_{Her_{(n=8)}}$	**4**	**4**
$KMD_{G_{(\sigma=0.001)}}$	**4**	**4**
$KMD_{P_{(a=2,\ b=2,\ d=10)}}$	**4**	**4**
KMD_S	**4**	**4**
$KMD_{Leg_{(n=8)}}$	**4**	**4**
$KMD_{Lag_{(n=8)}}$	**4**	**4**
$KMD_{Her_{(n=8)}}$	**4**	**4**
BD	2	4
MBD	0	4
MHRD	0	4
RTD	1	4
HMD	4	4

4.3 Real Data Experiment

Berkeley Growth Study data. We consider the Berkeley Growth Study data, that contain the heights of 39 boys and 54 girls from age 1 to 18 and the ages at which they were collected. First we consider all the boys (main data) and contaminate them with 5 randomly selected girls (the 'outlying' curves). This procedure was repeated 100 times so we obtain 100 random samples contaminated with outliers. Next we consider the opposite case, taking the girls as the main data contaminate them with groups of 5 randomly selected boys as 'outlying' data, see Fig. 8. The results are presented in table 3. Again the kernel depth obtains the best results in detecting the outliers in both cases.

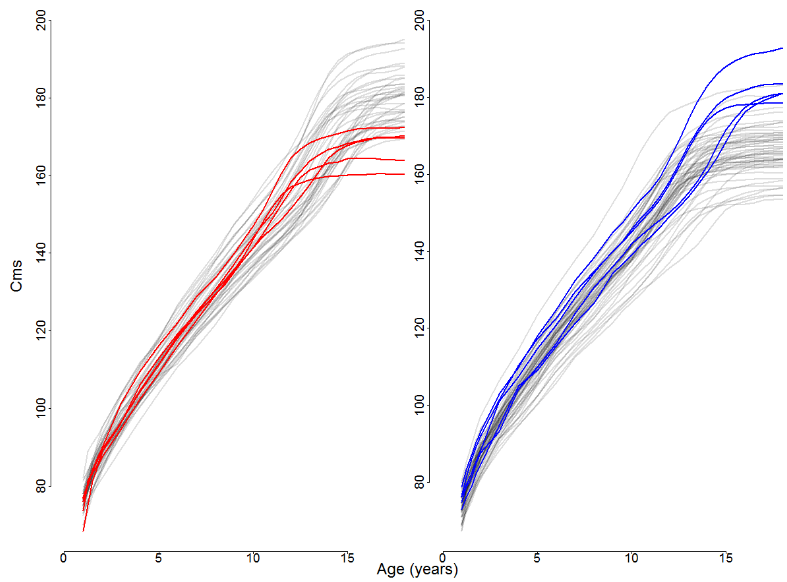

Figure 8: Left: Main population boys, contaminated by 5 girls. Right: Main population girls, contaminated by 5 boys.

Table 3: Mean and standard deviation (in parentheses) of the proportion of correctly identified outliers, for n=100 ($\gamma = 10^{-6}$).

Measures	Base data: Boys		Base data: Girls	
$GKD_{G_{(\sigma=0.014)}}$	**0,622**	(0,234)	**0,49**	(0,154)
$GKD_{P_{(a=2,\ b=2,\ d=10)}}$	0,572	(0,212)	0,308	(0,203)
GKD_S	0,618	(0,237)	0,456	(0,155)
$GKD_{Leg_{(n=8)}}$	0,572	(0,212)	0,308	(0,203)
$GKD_{Lag_{(n=8)}}$	0,572	(0,212)	0,308	(0,203)
$GKD_{Her_{(n=8)}}$	0,572	(0,212)	0,308	(0,203)
$KMD_{G_{(\sigma=0.014)}}$	0,512	(0,156)	0,352	(0,09)
$KMD_{P_{(a=2,\ b=2,\ d=10)}}$	0,513	(0,174)	0,36	(0,164)
KMD_S	0,486	(0,168)	0,345	(0,11)
$KMD_{Leg_{(n=3)}}$	0,513	(0,174)	0,36	(0,164)
$KMD_{Lag_{(n=3)}}$	0,513	(0,174)	0,36	(0,164)
$KMD_{Her_{(n=3)}}$	0,513	(0,174)	0,36	(0,164))
MBD	0,194	(0,166)	0,07	(0,095)
MHRD	0,195	(0,167)	0,07	(0,095)
RTD	0,216	(0,174)	0,144	(0,163)
HMD	0,39	(0,207)	0,194	(0,154)

Australia mortality rates. Here we consider age-specific mortality rates for Australian males for 1901-2003, in logarithmic scale, which is publicly available in the R package 'fds'. In this experiment we do not know a priori if there is an outlying curve, so we define as outlier the curve that satisfies that $Pr(KD < C) = 0.01$, where C is the inverse of the empirical distribution function of KD evaluated at $x = 0.01$, $C = F_{KD}^{-1}(x = 0.01)$, where KD is either the KMD or the GKD.

In a previous work Arribas-Gil and Romo (2014), the authors identifed a 'shape' outlier, corresponding to the mortality rate of the year 1919. The aim of this experiment is to demonstrate that kernel depths are also able to detect this type of outliers. The kernel parameters used are $\sigma = 0.01$ for the the Gaussian kernel and $a = 1,\ b = 1,\ d = 2$ for the polynomial kernel. The penalization coefficient of the SVM regulariztion problem is $\gamma = 0.015$ (except for the case of polynomial kernel where the γ considered was 1).

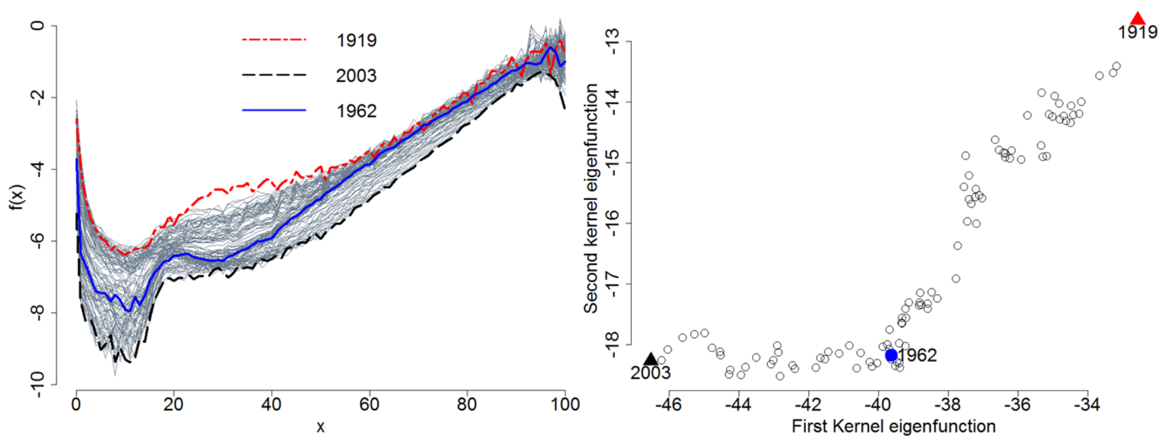

Figure 9: Observed curves and outliers detected. KMD and GKD outiliers: year 1919 in red (dash-dotted line), year 2003 in black (dashed line) and the deepest curve year 1962 (left panel). RKHS projections using the Gaussian kernel and outliers detected year 1919 (red triangle), year 2003 (black triangle) and the deepest point year 1962 (right panel).

The results presented in Figure 9 show that both kernel depths are able to identify the shape outlier (year 1919). The dash-dotted curve highlighted in red that correspond to the year 1919, is the shallowest curve. The dashed curve in black, that correspond to the year 2003, is the second most outlying curve. These outliers were identified using all kernel functions considered. The deepest curve in blue correspond to the year 1962. If we apply the competitor depth measures we find that the outliers detected are the mortality rate for the year 2003 and 2002. Both curves can be considered as extreme observations, but that share the same pattern with the rest of the curves of the sample (excepting the year 1919). The only competitor that detected the shape outlier (1919) was the HMD.

Vertical Density Profiles. The data set contains 24 curves of vertical density profiles (VDP). These data come from the manufacture of engineered woodboards. Figure 10-left shows the 24 profiles. Each one consists of 314 measurements taken 0.002 inches. In Moguerza et al. (2007), the authors identified 3 outliers profiles in red in Figure 10-right, in particular curves 3,6, and 16. The aim of this experiment is to show that the order induced by the the proposed depths functions allows us to identify these outliers in the sample of non-linear profiles. Both of our depth measures are able to identify the 3 non-linear profiles

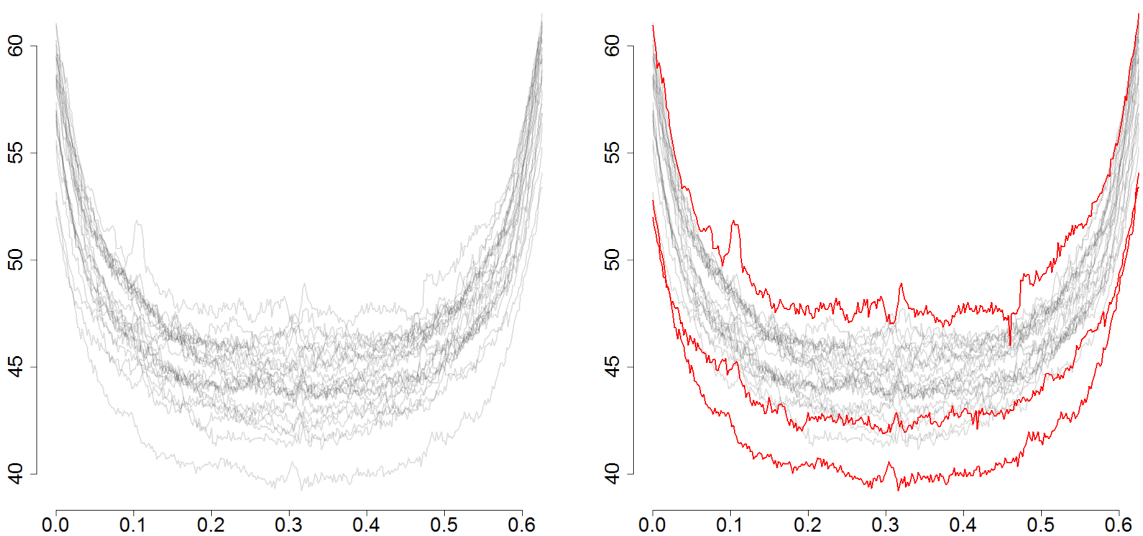

Figure 10: 24 profiles, each one consists of 314 measurements and the 3 otulier profiles in red (right panel).

idenfitifed as outliers highlighted in red in Figure 10. With respect to the competitor depth functions, only the MBD and MHRD are able to capture the shape outlier (curve 16). The remaining metrics, are able to detect two outliers which corresponds to curves 3 and 6 which are the most outlying ones, but not the shape outlier, (curve 16).

5 Conclusions

In this paper we introduce a new approach to handle order problems in functional data. In particular we present a new definition of deepest point for functional data that induces a center-outward ordering for functional data sets, and proposed a methodology to order functional data based on a representation of the functional datums via projections onto a Reproducting Kernel Hilbert Space. In this sense two depth measures that induce order into the data were proposed: i) the kernel Mahalanbobis depth, based on the Mahalanovis depth (Mahalanobis (1936)) and, ii) the generalized kernel depth based on a generalization of the Mahalanobis distance via density kernels proposed by Martos et al. (2013).

In propositions 3.2 to 3.5 we have shown that the generalized kernel depth satisfies the properties of *Affine invariance*, *Maximality at center*, *Monotonicity relative to the deepest point* and *Vanishing at infinity* proposed by Zuo and Serfling (2000). Moreover in propositions 3.1 we have shown that the h-mode depth proposed by Cuevas et al. (2007) is a particular case of the generalized kernel depth when $\phi_{\mathbb{P}}(\mathbf{x}) = \sum_{i=1}^{n} K_h(\|\mathbf{x} - c_i\|)$.

To complement the typical kernel functions used in the literature, we have constructed different kernels functions using orthogonal polynomials, such the Ledengre, Laguerre and Hermite polynomials as functions of a basis where the kernel can be constructed as $K(x, y) = \sum_j \lambda_j \phi_j(x)\phi_j(y)$. In Theorem 3.1 we have studied the indepedence of the generalized kernel depth with respect to the kernel choice.

In line with the above, we have shown that both kernel depths proposed improve the result of the most widely used depth measures for functional data. This results are presented for a different real data set and several sets of simulated data. Moreover some 'traditional' depth measures are based on plain representations of the curves which implies that the deepest curve is not necessarily invariant to affine transformations.

Regarding future work, we have presented a functional depth measure based on RKHS representation of the functional data. In this sense we propose to study the geometric properties of the projection space in order to analyze if the performance of ths statistical functional measure is preseved or enhanced when functional data lives in a non-linear space or manifold.

References

Arribas-Gil, A. and J. Romo (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics 15*(4), 603–619.

Chakraborty, A. and P. Chaudhuri (2014). On data depth in infinite dimensional spaces. *Annals of the Institute of Statistical Mathematics 66*(2), 303–324.

Cucker, F. and S. Smale (2002). On the mathematical foundations of learning. *Bulletin of the American Mathematical Society 39*, 1–49.

Cuesta-Albertos, J. A. and A. Nieto-Reyes (2008). The random tukey depth. *Computational Statistics & Data Analysis 52*(11), 4979–4988.

Cuevas, A., M. Febrero, and R. Fraiman (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics 22*(3), 481–496.

Fasshauer, G. and M. McCourt (2015). *Kernel-based approximation methods using Matlab*, Volume 19. World Scientific Publishing Co Inc.

Fraiman, R. and G. Muniz (2001). Trimmed means for functional data. *Test 10*(2), 419–440.

Liu, R. Y. et al. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics 18*(1), 405–414.

Liu, R. Y., J. M. Parelius, K. Singh, et al. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh). *The annals of statistics 27*(3), 783–858.

López-Pintado, S. and J. Romo (2009). On the concept of depth for functional data. *Journal of the American Statistical Association 104*(486), 718–734.

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

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta) 2*, 49–55.

Martos, G., A. Muñoz, and J. González (2013). On the generalization of the mahalanobis distance. In *Iberoamerican Congress on Pattern Recognition*, pp. 125–132. Springer.

Martos, G., A. Muñoz, and J. González (2014). Generalizing the mahalanobis distance via density kernels. *Intelligent Data Analysis 18*(6S), S19–S31.

Moguerza, J. M. and A. Muñoz (2006). Support vector machines with applications. *Statistical Science*, 322–336.

Moguerza, J. M., A. Muñoz, and S. Psarakis (2007). Monitoring nonlinear profiles using support vector machines. In *Iberoamerican Congress on Pattern Recognition*, pp. 574–583. Springer.

Muñoz, A. and J. González (2010). Representing functional data using support vector machines. *Pattern Recognition Letters 31*(6), 511–516.

Nagy, S. (2015). Consistency of -mode depth. *Journal of Statistical Planning and Inference 165*, 91 – 103.

Nieto-Reyes, A., H. Battey, et al. (2016). A topologically valid definition of depth for functional data. *Statistical Science 31*(1), 61–79.

Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science 72*, 1.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the international congress of mathematicians*, Volume 2, pp. 523–531.

Wahba, G. (1990). *Spline models for observational data.* SIAM.

Zhang, J., D. J. Olive, and P. Ye (2012). Robust covariance matrix estimation with canonical correlation analysis. *International Journal of Statistics and Probability 1*(2), 119.

Zuo, Y. and R. Serfling (2000). General notions of statistical depth function. *Annals of statistics*, 461–482.