UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 12-09
Statistics and Econometrics Series 06
May 2012

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

SPATIAL DEPTH-BASED CLASSIFICATION FOR FUNCTIONAL DATA

Carlo Sguera, Pedro Galeano and Rosa Lillo

Abstract

Functional data are becoming increasingly available and tractable because of the last technological advances. We enlarge the number of functional depths by defining two new depth functions for curves. Both depths are based on a spatial approach: the functional spatial depth (FSD), that shows an interesting connection with the functional extension of the notion of spatial quantiles, and the kernelized functional spatial depth (KFSD), which is useful for studying functional samples that require an analysis at a local level. Afterwards, we consider supervised functional classification problems, and in particular we focus on cases in which the samples may contain outlying curves. For these situations, some robust methods based on the use of functional depths are available. By means of a simulation study, we show how FSD and KFSD perform as depth functions for these depth-based methods. The results indicate that a spatial depth-based classification approach may result helpful when the datasets are contaminated, and that in general it is stable and satisfactory if compared with a benchmark procedure such as the functional *k*-nearest neighbor classifier. Finally, we also illustrate our approach with a real dataset.

Keywords: Depth notion; Spatial functional depth; Supervised functional classification; Depth-based method; Outliers.

Universidad Carlos III de Madrid, Department of Statistics, Facultad de Ciencias Sociales y Jurídicas, Campus de Getafe, Madrid, Spain. E-mail addresses: csguera@est-econ.uc3m.es (Carlo Sguera), pedro.galeano@uc3m.es (Pedro Galeano) and lillo@est-econ.uc3m.es (Rosa Lillo) **Acknowledgements**: This research was partially supported by Spanish Ministry of Education and Science grant 2007/04438/001, by Madrid Region grant 2011/00068/001, by Spanish Ministry of Science and Innovation grant 2012/00084/001 and by MCI grant MTM2008-03010.

Spatial Depth-Based Classification for Functional Data

Carlo Sguera
Department of Statistics, Universidad Carlos III de Madrid
28903 Getafe (Madrid), Spain
(csguera@est-econ.uc3m.es)
Pedro Galeano
Department of Statistics, Universidad Carlos III de Madrid
28903 Getafe (Madrid), Spain
(pedro.galeano@uc3m.es)
and
Rosa Lillo
Department of Statistics, Universidad Carlos III de Madrid
28903 Getafe (Madrid), Spain
(lillo@est-econ.uc3m.es)

Abstract

Functional data are becoming increasingly available and tractable because of the last technological advances. We enlarge the number of functional depths by defining two new depth functions for curves. Both depths are based on a spatial approach: the functional spatial depth (FSD), that shows an interesting connection with the functional extension of the notion of spatial quantiles, and the kernelized functional spatial depth (KFSD), which is useful for studying functional samples that require an analysis at a local level. Afterwards, we consider supervised functional classification problems, and in particular we focus on cases in which the samples may contain outlying curves. For these situations, some robust methods based on the use of functional depths are available. By means of a simulation study, we show how FSD and KFSD perform as depth functions for these depth-based methods. The results indicate that a spatial depth-based classification approach may result helpful when the datasets are contaminated, and that in general it is stable and satisfactory if compared with a benchmark procedure such as the functional k-nearest neighbor classifier. Finally, we also illustrate our approach with a real dataset.

Keywords: Depth notion; Spatial functional depth; Supervised functional classification; Depth-based method; Outliers.

1 INTRODUCTION

The technological advances of the last decades in many fields such as chemometrics, medicine, engineering or finance have allowed to observe and study random samples that are composed by curves. In these cases, although we observe a finite number of values for each curve, it is common to assume that the sample generating process consists in a stochastic function, that is to say a random variable taking values on an infinite-dimensional space. To analyze this type of data, it is convenient to use the tools provided by a quite new area of statistics known as functional data analysis (FDA). For two complementary FDA overviews, one parametric and the other nonparametric, see Ramsay and Silverman (2005) and Ferraty and Vieu (2006), respectively. The reasons why a standard multivariate data analysis might fail when data are curves are mainly three. First, although curves are usually observed as vectors, the evaluation points may differ in number and/or position from curve to curve and, unlike multivariate observations, they cannot be permuted. Second, but possibly more important, the data generating stochastic functions are characterized by a dependence structure. This feature brings about a considerable autocorrelation in functional observations and has repercussions on any attempt to analyze the data with standard multivariate procedures. Third, functional samples may contain less curves than evaluation points, and this aspect usually represents a great problem in multivariate data analysis.

However, the contraposition between multivariate and functional statistics is not total. As a matter of fact, many multivariate techniques have inspired advances in FDA, and the introduction of the notion of data depth for functional data represents an example. With the goal of measuring the degree of centrality of points with respect to probability distributions or random samples, a general notion of data depth is first born in the multivariate framework. The core of this notion has soon interested many authors working with functional data, who have proposed different functional implementations of the idea of depth (see Fraiman and Muniz 2001; Cuevas, Febrero, and Fraiman 2006; Cuesta-Albertos and Nieto-Reyes 2008; Cuevas and Fraiman 2009; López-Pintado and Romo 2009, among others). In this article, we present two new functional data depths which have as starting point a certain way to define quantiles in multivariate and functional spaces.

In one dimension, the definition of order statistics is straightforward and naturally arises

from the intrinsic order on the real line; notwithstanding their simplicity, order statistics have a great importance, especially in areas such as robust estimation and inference. In multidimensional spaces, there is no a unique natural order, but there is still interest in robust methods, and therefore in defining some notions of ordering. In this article, we will consider the proposal made by Chaudhuri (1996), who introduced a multivariate ordering criterion related to the notion of spatial quantiles, which are our main interest. Their definition can be extended to infinite-dimensional Hilbert and Banach spaces (Chaudhuri 1996) and they can be directly connected to a multivariate spatial depth function (Serfling 2002). Exploiting the previous two considerations, we propose the first contribution of this work, that is a spatial depth function for random elements belonging to infinite-dimensional Hilbert spaces, which we name functional spatial depth (FSD).

Similarly as observed by Chen, Dang, Peng, and Bart (2009) for the multivariate spatial depth proposed by Serfling (2002), we show that behind our FSD definition there is a global approach to the depth problem. In fact, if we consider a sample of curves and a curve x, all the sample observations will equally contribute to $FSD(x)$. However, for some samples might be interesting to study more carefully the neighborhoods of the curves, and therefore to dispose of a functional spatial depth coherent with a local approach. The second functional depth that we propose in this article will allow to study these samples, and it is the kernelized functional spatial depth (KFSD). In this case, the contribution of each sample observation to the value of the KFSD function at x will depend on the distance between each curve and x. More precisely, with KFSD we focus our analysis on the neighborhood of x, and neighboring curves will contribute more to $KFSD(x)$ than distant curves.

Both FSD and KFSD, as well as any functional depth, can be useful to perform exploratory FDA and to build robust functional methods. In this article, we tackle the supervised functional classification problem by considering three depth-based procedures that have been proposed by López-Pintado and Romo (2006) and Cuevas, Febrero, and Fraiman (2007). The depth-based classification methods were developed with robustness as main goal and have been mainly proposed for functional datasets that are possibly affected by the presence of some outlying curve. Actually, robustness might be a key issue in many functional classification problems, because the available FDA outlier detection procedures are still few (see for example Febrero, Galeano, and González-Manteiga 2008). For this rea-

son, the depth-based methods, through different decision rules, try to classify curves in a robust way: two of them by using the depth information provided by the training curves (López-Pintado and Romo 2006); the other by looking at the depth values of the curves to classify (Cuevas et al. 2007).

Nowadays, the depth-based procedures compete with a wide variety of supervised functional classification methods. For instance, Hastie, Buja, and Tibshirani (1995) proposed a penalized version of the multivariate linear discriminant analysis technique, whereas James and Hastie (2001) have directly built a functional linear discriminant analysis procedure that uses natural cubic spline functions to model the observations. Marx and Eilers (1999) organized functional supervised classification as a case of a generalized linear regression model, and proposed to use a P-spline approach. Hall, Poskitt, and Presnell (2001) suggested to perform dimension reduction by means of functional principal component analysis and to solve the derived multivariate problem with quadratic discriminant analysis or kernel methods. Ferraty and Vieu (2003) developed a functional kernel-type classifier. Epifanio (2008) proposed to describe curves by means of shape feature vectors and to use some classical multivariate classifier for the discrimination stage. Finally, Biau, Bunea, and Wegkamp (2005) and Cérou and Guyader (2006) both studied some consistency properties of the extension of the k-nearest neighbor procedure to infinite-dimensional spaces: the first extension considers a reduction of the dimensionality of the regressors based on a Fourier basis system; the second generalization deals with the real infinite dimension of the spaces under consideration.

Accordingly, our third contribution consists in studying how the proposed FSD and KFSD, and other existing functional depths, perform when used as the depth functions of the above-mentioned depth-based classification procedures. We also consider a benchmark, and in particular the version of the k-nearest neighbor method studied by Cérou and Guyader (2006). Our study considers many classification scenarios: initially, noncontaminated simulated functional data, and next contaminated simulated functional data, which represent indeed our main interest. The study shows that the spatial depth-based approach leads to good results, especially when the data are contaminated and when KFSD is the chosen spatial depth. Finally, we also classify curves belonging to a real dataset that is potentially characterized by the presence of outliers, and we obtain results which confirm the ones observed in the simulation study.

The article is organized as follows. The general supervised functional classification problem is introduced in Section 2, where we also present the existing depth-based methods and some functional depths. In Section 3 we define the functional spatial depth and the kernelized functional spatial depth. The results of our simulation and real data classification studies are presented in Sections 4 and 5, respectively. Finally, our conclusions are drawn in Section 6.

2 DEPTH-BASED SUPERVISED CLASSIFICATION FOR FUNCTIONAL DATA

A random variable Y is called functional variable if takes values in a functional space. An observation of Y is called functional data, and therefore a functional dataset consists in the observation of n functional variables identically distributed as Y (Ferraty and Vieu 2006). In what follows, let the functional space to be a Hilbert space $\mathbb{H}$ and $\|\cdot\|$ to be the norm defined by the inner product on $\mathbb{H}$. In practice, the elements of $\mathbb{H}$ are observed in form of curves at a discretized and finite set of different domain points $t_1 < t_2 < \ldots < t_{m-1} < t_m$, and the sets can be different from one curve to another.

In supervised functional classification, the natural theoretical framework is given by the random pair (Y, G), where Y is a functional random variable and G is a categorical random variable describing the class membership of each observation. For simplicity, G is usually a random variable with value $g = 0$ or $g = 1$, and in the rest of the article let this assumption to hold. Assume to observe a sample of n independent pairs, $(y_i, g_i)_{i=1,\ldots,n}$, identically distributed as (Y, G), and an independent curve, x, identically distributed as Y, but with unknown class membership. Using the information contained in the observed pairs $(y_i, g_i)_{i=1,\ldots,n}$, any supervised functional classification method provides a classification rule which can be used to classify the curve x (Ferraty and Vieu 2006).

The k-nearest neighbor procedure (k-NN) is one of the most popular methods used to perform supervised functional classification. Its generalization to infinite-dimensional spaces has been studied by Cérou and Guyader (2006) and it consists in the following rule: look

at the k nearest neighbors of x among $(y_i)_{i=1,\dots,n}$, and choose 0 or 1 for its label according to the majority vote. Under our assumptions, i.e. $(y_i)_{i=1,\dots,n}$ and $x \in \mathbb{H}$, the search of the neighbors is based on the norm defined on $\mathbb{H}$. Classification procedures such as k-NN, or the ones mentioned in the Introduction, might suffer from the presence of outliers in the data, which are also difficult to detect. Thus, the availability of robust functional classification techniques is crucial.

In univariate statistics, when exists the possibility that the data may contain outliers, a natural way to acquire robustness consists in considering methods based on order statistics. Clearly, tools such as order statistics are not naturally available in $\mathbb{R}^d$ and in functional spaces, but it is still of interest to dispose of some criterion to order multivariate observations and curves. One possibility consists in considering some center-outward ordering criterion, such as for example the one based on the general notion of data depth.

The notion of data depth originated in nonparametric multivariate analysis. According to Serfling (2006), a multivariate depth is a function which provides a P-based center-outward ordering of points $\mathbf{x} \in \mathbb{R}^d$, where P is a probability distribution on $\mathbb{R}^d$. Hence, the values of a multivariate depth function should be high at points that are central with respect to the probability distribution P, and low at peripheral points. This notion can be extended to more general settings, and in particular to functional spaces.

Several functional data depths have been proposed by various authors, and in what follows we recall briefly five notions. Fraiman and Muniz (2001) defined the Fraiman and Muniz depth (FMD), which tries "to measure how long remains a curve in the middle of a group of them"(Fraiman and Muniz 2001, p. 421). Based on the idea of measuring how densely a curve is surrounded by other curves of the sample, Cuevas et al. (2006) proposed the h-modal depth (HMD), which consists in a kernel-based function involving norms. Obviously, this depth depends on the choices about the norm and the kernel functions. To a completely different class belong the random Tukey depth (RTD, Cuesta-Albertos and Nieto-Reyes 2008) and the integrated dual depth (IDD, Cuevas and Fraiman 2009). Both are based on p random one-dimensional projections of the curves, which at the end generate p-dimensional multivariate points. The main difference between the two proposals is how these vectors are managed: the RTD function makes use of the multivariate Tukey depth; the IDD function makes use of the multivariate simplicial depth. Finally, López-Pintado and

Romo (2009) proposed a depth function based on the graphic representation of the curves, more precisely on all the possible bands defined by the graphs on the plane of $2, 3, \ldots$ and J curves, and on a measure of the sets where another curve is inside these bands. This depth is known as the modified band depth (MBD).

The above-mentioned existing functional depths, as well as the functional spatial depths that we introduce in Section 3, can be used to perform supervised functional classification together with some depth-based method. Before presenting the functional spatial depths, let us describe briefly the three depth-based methods that we consider in our simulation and real data studies. For all of them, assume to observe a sample of $n = n_0 + n_1$ independent pairs, identically distributed as (Y, G), where n_0 and n_1 are the sample sizes of the groups, and to observe an independent curve x, identically distributed as Y, but with unknown class membership. The first procedure is known as the distance to the trimmed mean method (DTM, López-Pintado and Romo 2006): for each of the two groups, DTM computes the α-trimmed mean m_g^α and classifies x in the group for which $\|x - m_g^\alpha\|$ is less. Clearly, the contribution of the functional depth will be at the trimming stage, allowing to obtain a robust mean. The second procedure is known as the weighted averaged distance method (WAD, López-Pintado and Romo 2006): for a given group, say the group with curves having label equal to 0, WAD computes a weighted average of the distances $\|x - y_i\|_{i=1,\ldots,n_0}$, where the weights are given by the within-group depth values $D(y_i)_{i=1,\ldots,n_0}$. WAD classifies x in the group for which the weighted averaged distance is less. The third procedure can be defined as the within maximum depth method (WMD, Cuevas et al. 2007): for a given group, say the group with curves having label equal to 0, WMD includes the curve x in the sample and computes its depth value, $D(x; g = 0)$. WMD classifies x in the group for which $D(x; \cdot)$ is higher.

3 FUNCTIONAL SPATIAL DEPTHS

In this section we present two new functional data depths, both defined considering as starting point the general idea of spatial depth. The origins of the spatial approach date back to Brown (1983), who studied the problem of robust location estimation in two-dimensional spatial data and introduced the idea of spatial median. This approach consists in consid-

ering the geometry of bivariate data and has been extended to $\mathbb{R}^d$ to provide the notion of multivariate spatial quantiles (Chaudhuri 1996) and a multivariate spatial depth function (Serfling 2002), which is defined as follows: let $\mathbf{x} \in \mathbb{R}^d$ and let $S : \mathbb{R}^d \to \mathbb{R}^d$ to be the multivariate spatial sign function given by

$$S(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{\|\mathbf{x}\|_E}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases} \tag{1}$$

where $\|\mathbf{x}\|_E$ is the Euclidean norm of $\mathbf{x}$. Let $\mathbf{Y}$ to be a random variable with cumulative distribution function F on $\mathbb{R}^d$. Then, the spatial depth of $\mathbf{x}$ with respect to F is defined by

$$SD(\mathbf{x}, F) = 1 - \left\| \int S(\mathbf{x} - \mathbf{y})\, dF(\mathbf{y}) \right\|_E = 1 - \|\mathbb{E}\left[S(\mathbf{x} - \mathbf{Y})\right]\|_E, \tag{2}$$

Note that (1) and (2) are practically two particular cases of two more general definitions: the first, the definition of a spatial sign function for elements belonging to normed vector spaces; the second, the definition of a spatial depth function for random elements belonging to normed vector spaces. In this article, we consider these two general definitions, but we focus on a different application for each of them. In more detail, we define the functional spatial sign function and the functional spatial depth function as

$$FS(x) = \begin{cases} \frac{x}{\|x\|}, & x \neq 0, \\ 0, & x = 0, \end{cases} \tag{3}$$

and

$$FSD(x, P) = 1 - \|\mathbb{E}\left[FS(x - Y)\right]\|, \tag{4}$$

where now $x \in \mathbb{H}$ and Y is a random variable with probability distribution P on $\mathbb{H}$. When a sample of curves is observed, i.e. $(y_i)_{i=1,\ldots,n}$, (4) must be replaced to compute the depth value of x with respect to the observed sample. To do this, we define the sample FSD as

$$FSD_n(x) = 1 - \frac{1}{n} \left\| \sum_{y \in (y_i)_{i=1,\ldots,n}} FS(x - y) \right\|. \tag{5}$$

Apart from defining the multivariate spatial depth, Serfling (2002) has also showed that

there is a connection between the notions of spatial depth and quantiles. In $\mathbb{R}^d$, $Q_F(\mathbf{u})$ is the $\mathbf{u}$th spatial quantile of the random variable $\mathbf{Y}$ with cumulative distribution function F if and only if $Q_F(\mathbf{u})$ is the value of $\mathbf{q}$ which minimizes

$$\mathbb{E}\left[\Phi(\mathbf{u}, \mathbf{Y} - \mathbf{q}) - \Phi(\mathbf{u}, \mathbf{Y})\right], \tag{6}$$

where $\Phi(\mathbf{u}, \mathbf{v}) = \|\mathbf{v}\|_E + \langle \mathbf{u}, \mathbf{v} \rangle_E$, $\langle \mathbf{u}, \mathbf{v} \rangle_E$ is the Euclidean inner product of $\mathbf{u}$ and $\mathbf{v}$, and $\{\mathbf{u} \colon \mathbf{u} \in \mathbb{R}^d, \|\mathbf{u}\|_E < 1\}$. Note that the $\mathbf{u}$th spatial quantile $Q_F(\mathbf{u})$ is characterized by the direction and the magnitude of $\mathbf{u}$. If the cumulative distribution function F is not supported on a straight line, $Q_F(\mathbf{u})$ is unique, and it is possible to define a spatial quantile function Q_F which associates to each $\mathbf{u}$ a unique element in $\mathbb{R}^d$. Moreover, there exists another way to characterize the $\mathbf{u}$th spatial quantile. In fact, $Q_F(\mathbf{u})$ can be represented as the solution of the following equation:

$$-\mathbb{E}\left[\frac{\mathbf{Y} - \mathbf{q}}{\|\mathbf{Y} - \mathbf{q}\|_E}\right] = \mathbf{u}. \tag{7}$$

Let $\mathbf{x}$ to be the solution of (7), then the left-hand side can be interpreted as the inverse of Q_F at $\mathbf{x}$; that is, $Q_F^{-1}(\mathbf{x}) = -\mathbb{E}\left[(\mathbf{Y} - \mathbf{x})/\|\mathbf{Y} - \mathbf{x}\|\right] = \mathbf{u}$. And finally, considering Euclidean norms, we have that

$$\|Q_F^{-1}(\mathbf{x})\|_E = \left\| -\mathbb{E}\left[\frac{\mathbf{Y} - \mathbf{x}}{\|\mathbf{Y} - \mathbf{x}\|}\right] \right\|_E = \|\mathbb{E}\left[S(\mathbf{x} - \mathbf{Y})\right]\|_E = 1 - SD(\mathbf{x}, F), \tag{8}$$

which shows the direct connection between the multivariate notions of spatial depth and quantiles (Serfling 2002).

Moving from $\mathbb{R}^d$ to an infinite Hilbert space, this connection still holds. Similarly as in the multivariate case, the uth functional spatial quantile of a $\mathbb{H}$-valued random variable Y with probability distribution P is obtained by minimizing with respect to q

$$\mathbb{E}\left[\Phi(u, Y - q) - \Phi(u, Y)\right], \tag{9}$$

where $\Phi(u, v) = \|v\| + \langle u, v \rangle$, $\langle u, v \rangle$ is the inner product of u and v, and $\{u \colon u \in \mathbb{H}, \|u\| < 1\}$. When Y is not concentrated on a straight line and it is not strongly concentrated around single points, Cardot, Cénac, and Zitt (2011) showed that the Fréchet derivative of the

convex function in (9) is given by

$$\Phi(q) = -\mathbb{E}\left[\frac{Y-q}{\|Y-q\|}\right] - u, \tag{10}$$

and that the uth quantile of Y is given by the unique solution of the equation $\Phi(q) = 0$. Let x to be the solution of $\Phi(q) = 0$, then the previous results allow to define the functional spatial quantile function FQ_P and its inverse FQ_P^{-1}. Evaluating FQ_P^{-1} at x, and considering norms, we have that

$$\|FQ_P^{-1}(x)\| = \left\|-\mathbb{E}\left[\frac{Y-x}{\|Y-x\|}\right]\right\| = \|\mathbb{E}\left[FS(x-Y)\right]\| = 1 - FSD(x,P), \tag{11}$$

which shows that the direct connection between the notions of spatial depth and quantiles also holds in Hilbert spaces. Therefore, the result in (11) enriches FSD with an interesting interpretability property and lays the foundations for a further theoretical study of FSD.

To introduce the second functional spatial depth, let us start with a comment about $FSD_n(x)$ defined in (5). Recall that for any $y \in (y_i)_{i=1,\dots,n}$ the functional spatial sign function value $FS(x-y)$ represents a unit-norm curve that is interpretable as the direction from x to y. Therefore, $FSD_n(x)$ depends on the sum of these n directions, and any sample observation contributes equally to $FSD_n(x)$ through the unit-norm curve $FS(x-y)$. This feature, the equal contribution of the sample observations to $FSD_n(x)$, is an important property, but generates a trade-off: on one side, it makes the FSD function robust; on the other, it transforms $(x-y)$ into a unit-norm curve regardless of y is a neighboring or a distant curves from x. Similarly as observed by Chen et al. (2009) for the multivariate spatial depth, this feature of FSD is due to a global approach to the problem, but in some circumstances a more local approach might be of interest. Indeed, with the implementation of a local approach, the information brought by each observation would depend on its distance from x, whereas with a global approach each observation brings the same amount of information. In Figure 1 we show two illustrative cases in which some of the depths presented in Section 2 (FMD, RTD, IDD, MBD) and the depth introduced in this section (FSD) fail in a big way because they are based on a global approach. The examples clearly involve three strongly different classes of curves, but we consider them as belonging to an homogeneous sample. Despite this point, the examples still serve to highlight a serious drawback of what we are

describing as a global approach. In the first example, we generate 21 curves from a given process and we divide them in three groups with 10, 10 and 1 curves, respectively. To each group of curves, we add a different constant (0, 10 and 5, respectively) and we compute the depth values of the transformed curves: for all the global-oriented depths, the dashed curve results as the deepest curve.

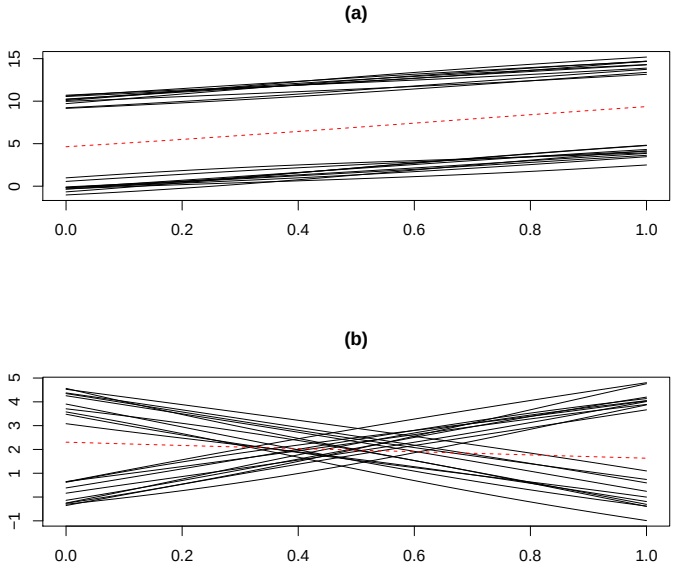

Figure 1: Two illustrative examples: the global-oriented depths attain their maximum values at the dashed curves; the local-oriented depths attain their minimum values at the dashed curves.

The structure of the second example is similar to the first one, but we use a different kind of transformation for the curves belonging to the second and third group; also in this case, we observe that the dashed curve is the deepest curve for all the global-oriented depths. Evidently, even if in both examples there is little interest in considering the curves as a one-class sample, in presence of nonunimodal data the behaviors of the global-oriented depths turn out strongly inconvenient. Indeed, it would be more reasonable to observe at least low depth values for the dashed curves, as occurs for HMD and for the local-oriented version of FSD that we introduce hereafter. For both these depths the dashed curves result as the less deep curves.

A classical way to implement a local approach consists in considering some kernel-based method. As in the multivariate case, it is possible to show that

$$\left\| \sum_{y \in (y_i)_{i=1,\ldots,n}} FS(x-y) \right\|^2 =$$
$$= \sum_{y,z \in (y_i)_{i=1,\ldots,n}} \frac{\langle x,x\rangle + \langle y,z\rangle - \langle x,y\rangle - \langle x,z\rangle}{\sqrt{\langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle}\sqrt{\langle x,x\rangle + \langle z,z\rangle - 2\langle x,z\rangle}}. \tag{12}$$

The right-hand side of (12) involves inner products, which can also be seen as similarity measures. The idea of Chen et al. (2009) consists in recoding the data to obtain a more powerful similarity measure. Their approach is common in areas such as machine learning and pattern analysis, and it is usually implemented by considering a positive definite and stationary kernel function instead of the inner product function; that is,

$$\kappa(x,y) = \langle \phi(x), \phi(y) \rangle. \tag{13}$$

In our case, x and $y \in \mathbb{H}$, and $\phi : x \in \mathbb{H} \to \mathbb{F}$ is an embedding map. Note that the map ϕ and the space $\mathbb{F}$ are usually defined implicitly by choosing or defining a kernel function. Then, exploiting (12) and substituting the inner product by the kernel introduced in (13), we define the sample kernelized functional spatial depth,

$$KFSD_n(x) = 1 -$$
$$\frac{1}{n}\left(\sum_{y,z \in (y_i)_{i=1,\ldots,n}} \frac{\kappa(x,x) + \kappa(y,z) - \kappa(x,y) - \kappa(x,z)}{\sqrt{\kappa(x,x) + \kappa(y,y) - 2\kappa(x,y)}\sqrt{\kappa(x,x) + \kappa(z,z) - 2\kappa(x,z)}} \right)^{1/2}, \tag{14}$$

which can also be interpreted as a recoded version of $FSD_n(x)$, that is

$$KFSD_n(x) = 1 - \frac{1}{n}\left\| \sum_{\phi(y) \in (\phi(y_i))_{i=1,\ldots,n}} FS(\phi(x) - \phi(y)) \right\| = FSD_{n_\phi}(\phi(x)), \tag{15}$$

where n_ϕ indicates that the original sample is replaced by the recoded sample. Note that the same idea applies to the other existing kernel-based depth, HMD (Cuevas et al. 2006). The sample version of HMD is given by

$$HMD_n(x) = \sum_{y \in (y_i)_{i=1,\dots,n}} \kappa(x, y), \tag{16}$$

and if we define the inner products sum function as

$$IPS_n(x) = \sum_{y \in (y_i)_{i=1,\dots,n}} \langle x, y \rangle, \tag{17}$$

$HMD_n(x)$ can be interpreted as a recoded version of $IPS_n(x)$, that is

$$HMD_n(x) = \sum_{\phi(y) \in (\phi(y_i))_{i=1,\dots,n}} \langle \phi(x), \phi(y) \rangle = HMD_{n_\phi}(\phi(x)). \tag{18}$$

Going back to (15), we can also implicitly define the population kernelized functional spatial depth, that is

$$KFSD(x, P) = 1 - \|\mathbb{E}\left[FS(\phi(x) - \phi(Y))\right]\| = FSD(\phi(x), P_\phi), \tag{19}$$

where $\phi(Y)$ and P_ϕ are a recoded version of Y and P, respectively.

After introducing FSD and KFSD, in Sections 4 and 5 we study how they can be useful in functional supervised classification problems. At this phase, we do not investigate how the choices about κ or the kernel bandwidth would affect the behavior of KFSD, neither we consider how different assumptions about the functional space, and so about the norm function, would affect FSD and KFSD. We leave the study of these important aspects for the next stages of our research.

4 SIMULATION STUDY

In Sections 2 and 3 we presented three different depth-based procedures (DTM, WAD and WMD) and seven different data depths (FMD, HMD, RTD, IDD, MBD, FSD and KFSD). Pairing all the procedures with all the depths, we obtain 21 depth-based methods, and we want to compare them in terms of functional supervised classification capabilities (from now on, we refer to each method by the notation depth-based procedure+data depth: for example, DTM+FMD refers to the method obtained by using the DTM procedure together with the

FMD). Both methods and depths may depend on some parameters or assumptions. As regards the methods, DTM is the only one which depends on a parameter, and in particular on a trimming parameter α, that we set at $\alpha = 0.2$. On the other hand, with regard to the data depths, we make use of the next parameters and assumptions: for HMD, we follow the recommendations made by Febrero et al. (2008) in their work about depth-based functional outlier detection, which are L^2 norm, $\kappa(x, y) = (2/\sqrt{2\pi}) \times \exp(-\|x - y\|^2/2\sigma^2)$ and the bandwidth σ equal to the 15th percentile of the empirical distribution of $\{\|y_i - y_j\|, i, j = 1, \ldots, n\}$. In addition, to make the range of HMD equal to $[0, 1]$, we consider a normalized version of HMD. For RTD and IDD, the number of random projections is $p = 50$ and we generate the random directions through a Gaussian process. For MBD, the maximum number of curves defining a band is $J = 2$. For FSD, we assume that the observations belong to the $L^2[0, 1]$ space. We make the same functional space assumption for KFSD, for which we use a Gaussian kernel, in particular a functional version of the one proposed by Chen et al. (2009),

$$\kappa(x, y) = \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right), \tag{20}$$

and the same bandwidth as in HMD. Finally, as regards the benchmark procedure, we take into account $k = 5$ nearest neighbors.

Our simulation study is partially based on the ones performed by López-Pintado and Romo (2006) and by Cuevas et al. (2007), but it contains some slight differences. It can be divided in two parts: in the first one, we do not allow for contaminated data, whereas in the second one we allow for them through a contamination probability given by q.

In absence of contamination, the curves generating processes have the following common structure:

$$x(t) = m_g(t) + \epsilon(t), \quad t \in [0, 1], \tag{21}$$

where $m_g(t)$ is a deterministic mean function characterizing the group g and $\epsilon(t)$ is a zero-mean Gaussian component. We consider two-groups problems ($g = 0$ or $g = 1$), and two different scenarios for the pair of mean functions, $m_0(t)$ and $m_1(t)$:

- PM1 (pair of means 1): $m_0(t) = 4t; \quad m_1(t) = 7t.$
- PM2 (pair of means 2): $m_0(t) = 15(1-t)t^{1.5}; \quad m_1(t) = 15(1-t)^{1.5}t.$

Note that PM1 is composed by linear functions, whereas PM2 by nonlinear functions. In addition, through different covariance functions for $\epsilon(t)$, we consider two different dependence structures in the simulated functional data:

- DS1 (dependence structure 1): $\mathbb{E}(\epsilon(t), \epsilon(s)) = 0.25 \exp\{-(t-s)^2\}, \quad t, s \in [0, 1].$
- DS2 (dependence structure 2): $\mathbb{E}(\epsilon(t), \epsilon(s)) = 0.3 \exp\{-|t-s|/0.3\}, \quad t, s \in [0, 1].$

Note that DS1 implies a stronger dependence than DS2. In summary, when there is no contamination, combining the scenarios for the mean functions with the scenarios for the dependence structures, we get 4 different models. In Figure 2 we report a simulated dataset for each model.

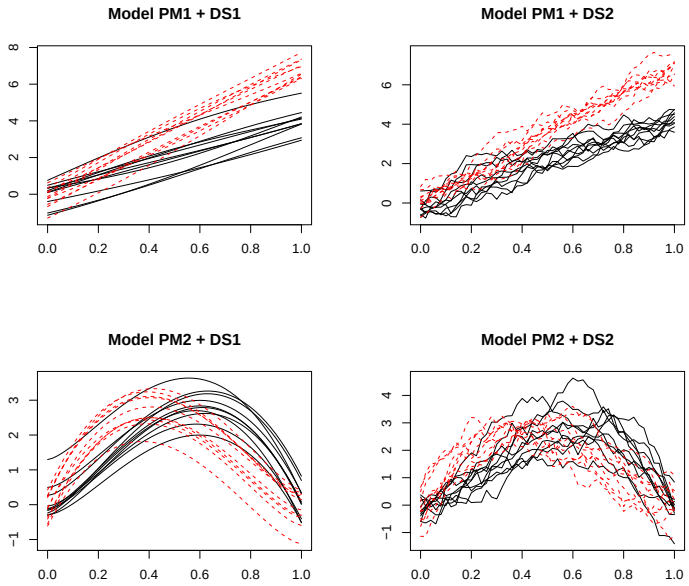

Figure 2: Simulated datasets from models PM1+DS1, PM1+DS2, PM2+DS1 and PM2+DS2: each dataset contains 10 curves from group $g = 0$ and 10 dashed curves from group $g = 1$.

When we allow for contamination, we consider four types of contamination for PM1 and one type of contamination for PM2. In all the five cases, the contamination affects only $m_1(t)$. We describe the five contamination scenarios in what follows.

- C1 (asymmetric total magnitude contamination for PM1):

$$m_0(t) = 4t; \quad m_1(t) = 7t + bM,$$

where $b \sim$ Bernoulli(q), $q = 0.1$ and $M = 5$.

- C2 (symmetric total magnitude contamination for PM1):

$$m_0(t) = 4t; \quad m_1(t) = 7t + bsM,$$

where s takes value 1 with probability equal to 1/2 and -1 with probability equal to 1/2.

- C3 (symmetric partial magnitude contamination for PM1):

$$m_0(t) = 4t; \quad m_1(t) = \begin{cases} 7t, & t < u \\ 7t + bsM, & t \geq u \end{cases},$$

where $u \sim$ Unif$[0, 1]$.

- C4 (asymmetric peaks magnitude contamination for PM1):

$$m_0(t) = 4t; \quad m_1(t) = \begin{cases} 7t, & t \notin [u, u + l] \\ 7t + bM, & t \in [u, u + l] \end{cases},$$

where $u \sim$ Unif$[0, 1 - l]$ and $l = 1/6$ is the fixed length of the peaks.

- C5 (shape contamination for PM2):

$$m_0(t) = 15(1 - t)t^{1.5}; \quad m_1(t) = 15(1 - t)^{1.5+bS}t,$$

where $S = 1$.

In presence of contamination, pairing C1, C2, C3, C4 and C5 with each of the dependence structures, we get 10 different models. In Figure 3 we report a simulated dataset for the models having dependence structure DS1.

The details of the simulation study are the following ones: for each model, we generate 125 replications. For each replication, we generate 100 curves, 50 for $g = 0$ and 50 for $g = 1$. We use 25 curves from $g = 0$ and 25 curves from $g = 1$ to build each training sample, and the remaining curves to build each test sample. All curves are generated using a discretized

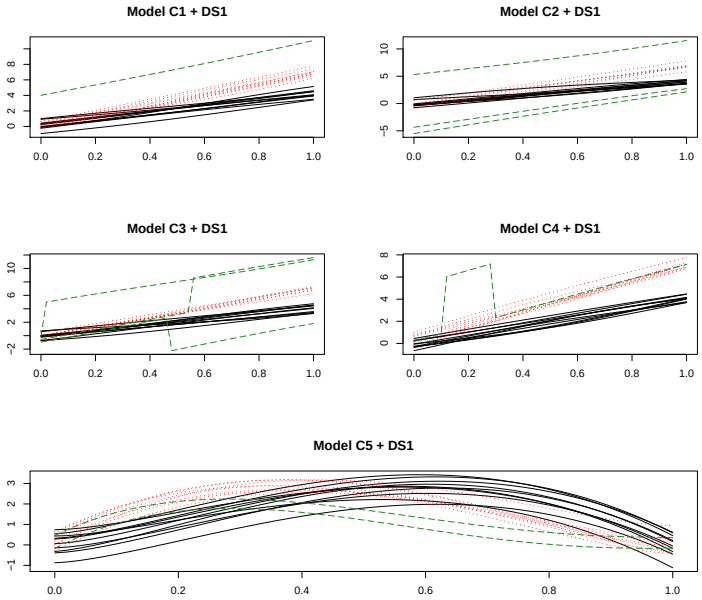

Figure 3: Simulated datasets with contamination of type C1, C2, C3, C4, C5 and dependent structure DS1: each dataset contains 10 curves from group $g = 0$ and 10 curves from group $g = 1$. The curves from group $g = 1$ can be noncontaminated (dotted) or contaminated (longdashed).

and finite set of 51 equidistant points between 0 and 1. For all the functional depths, we use a discretized version of their definition. We perform the comparison among methods in terms of misclassification percentages. Due to the large number of methods and models, we carry out the comparison by reporting in the next tables the means and the coefficients of variation of the misclassification percentages. In each table, we report the three lowest depth-based methods means in bold type.

Tables 1-4 report the performances observed when considering the models which do not allow for contamination. These first results show a clear but expected fact: in absence of contamination, k-NN outperforms the depth-based methods in terms of mean misclassification percentages. All the same, comparing carefully the depth-based methods and the benchmark performances, we also observe that:

1. WAD+KFSD is the unique depth-based method for which the mean misclassification percentages are never more than 2 times greater than the ones observed for the benchmark. In terms of variability, considering the standard deviations of the methods, k-NN still outperforms WAD+KFSD. However, taking into account the differences in

Table 1: PM1+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	1.66	1.81	1.66	1.79	1.74	1.63	1.78
	(1.16)	*(1.07)*	*(1.01)*	*(1.11)*	*(1.10)*	*(1.12)*	*(1.07)*
WAD	**1.41**	1.65	1.58	1.52	**1.50**	1.58	**1.38**
	(1.17)	*(1.11)*	*(1.07)*	*(1.13)*	*(1.10)*	*(1.13)*	*(1.11)*
WMD	10.50	2.18	11.04	3.92	8.66	3.12	1.58
	(1.17)	*(1.11)*	*(1.07)*	*(1.13)*	*(1.10)*	*(1.13)*	*(1.11)*
k-NN				0.69			
				(1.57)			

Table 2: PM1+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	0.35	0.37	**0.30**	0.35	**0.32**	0.34	0.35
	(2.29)	*(2.33)*	*(2.65)*	*(2.40)*	*(2.43)*	*(2.48)*	*(2.40)*
WAD	0.34	0.35	**0.29**	0.34	**0.32**	**0.32**	**0.32**
	(2.36)	*(2.40)*	*(2.60)*	*(2.36)*	*(2.43)*	*(2.43)*	*(2.43)*
WMD	3.20	1.95	17.46	0.96	2.18	0.35	0.59
	(0.77)	*(1.24)*	*(0.34)*	*(1.44)*	*(0.93)*	*(2.17)*	*(2.10)*
k-NN				0.30			
				(2.51)			

Table 3: PM2+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	2.06	2.30	2.00	2.00	1.98	1.90	2.34
	(1.36)	*(1.35)*	*(1.51)*	*(1.37)*	*(1.39)*	*(1.38)*	*(1.25)*
WAD	**1.38**	2.19	1.55	1.58	**1.44**	1.57	**1.44**
	(1.48)	*(1.29)*	*(1.52)*	*(1.45)*	*(1.49)*	*(1.43)*	*(1.34)*
WMD	25.14	3.12	12.42	23.47	21.04	7.04	2.03
	(0.26)	*(1.08)*	*(0.50)*	*(0.32)*	*(0.28)*	*(0.71)*	*(1.17)*
k-NN				0.88			
				(1.57)			

Table 4: PM2+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	6.96	6.94	6.91	**6.70**	6.99	6.86	6.99
	(0.56)	*(0.57)*	*(0.55)*	*(0.58)*	*(0.59)*	*(0.55)*	*(0.58)*
WAD	**6.66**	6.96	6.98	6.77	**6.70**	6.78	6.78
	(0.56)	*(0.58)*	*(0.57)*	*(0.54)*	*(0.57)*	*(0.57)*	*(0.57)*
WMD	13.20	8.58	20.45	21.82	11.81	7.49	**6.70**
	(0.38)	*(0.54)*	*(0.30)*	*(0.32)*	*(0.39)*	*(0.51)*	*(0.59)*
k-NN				6.59			
				(0.58)			

terms of means, and considering the coefficients of variation as appropriate variability measures, k-NN and WAD+KFSD show a similar variability.

2. Analyzing the supervised classification procedures behaviors, WAD results the best one: in 21 cases over 28 its means are less than 2 times greater than the benchmark means, and in one of these cases its mean is lower, that is WAD+RTD under model PM1+DS2. For DTM, the cases decrease to 14, whereas WMD behaves reasonably well only with KFSD. Therefore, for a depth-based supervised classification analysis characterized by the absence of outliers, the above considerations imply that WMD (except WMD+KFSD, and maybe WMD+HMD) should not be used to solve the problem, and that WAD should be preferred to DTM.
3. Analyzing the behaviors of the depths and looking at the percentages in bold type, we observe that FMD and MBD perform well when used together with the WAD procedure. As regards the spatial depths, FSD performs reasonably well together with WAD, but worst than its kernelized version, which indeed is the main competitor for FMD and MBD as the best depth for the WAD procedure. Moreover, as commented at point 2., KFSD is also the unique depth for which we observe an acceptable behavior in WMD.
4. The differences in terms of performances between the benchmark and the depth-based

methods shrink in presence of the weak dependence structure (DS2, Tables 2 and 4). This result may have an important implication: in presence of curves with weak dependence, when a pre-smoothing step is required or wanted, it should not hide this feature of the data to not compromise a depth-based supervised classification analysis.

Tables 5-14 report the performances observed when considering the 10 remaining models, the ones which allow for contamination. Also in this case the benchmark turns out to be an extremely competitive procedure, which means that k-NN performs quite well even with outliers. However, we observe that in 4 situations there is at least a depth-based method which outperforms k-NN, and we refer to the 2 models allowing for symmetric total magnitude contamination (C2+DS1 and C2+DS2), and to the models allowing for symmetric partial magnitude and shape contamination, and having DS2 (C3+DS2 and C5+DS2). For all these cases, the best depth-based method involves always KFSD (WMD+KFSD for C2+DS1, C2+DS2 and C3+DS2; WAD+KFSD for C5+DS2). Moreover, we also observe that:

Table 5: C1+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	1.73 *(1.56)*	**1.58** *(1.67)*	2.54 *(1.24)*	2.02 *(1.46)*	2.05 *(1.44)*	2.06 *(1.42)*	1.79 *(1.84)*
WAD	2.34 *(1.37)*	**1.34** *(1.44)*	1.70 *(1.24)*	1.87 *(1.29)*	1.95 *(1.30)*	**1.62** *(1.28)*	2.66 *(1.32)*
WMD	10.62 *(0.44)*	3.18 *(1.01)*	9.84 *(0.53)*	3.89 *(0.79)*	8.90 *(0.50)*	3.12 *(0.90)*	2.42 *(1.14)*
k-NN				0.45 *(2.26)*			

Table 6: C1+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	**0.46** *(1.91)*	**0.48** *(2.01)*	1.68 *(1.57)*	0.61 *(2.06)*	0.54 *(1.95)*	0.51 *(2.04)*	0.56 *(2.01)*
WAD	2.06 *(1.38)*	**0.46** *(1.91)*	0.77 *(1.80)*	1.10 *(1.60)*	1.14 *(1.59)*	0.82 *(1.76)*	2.54 *(1.35)*
WMD	4.22 *(0.70)*	3.10 *(1.00)*	13.57 *(0.43)*	1.58 *(1.04)*	3.34 *(0.80)*	0.75 *(1.50)*	1.31 *(1.54)*
k-NN				0.40 *(2.20)*			

Table 7: C2+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	4.51 *(0.65)*	4.50 *(0.64)*	4.35 *(0.63)*	4.48 *(0.61)*	4.40 *(0.65)*	4.43 *(0.64)*	4.51 *(0.65)*
WAD	4.88 *(0.59)*	4.22 *(0.63)*	**4.19** *(0.63)*	4.32 *(0.65)*	4.38 *(0.64)*	4.30 *(0.64)*	5.26 *(0.62)*
WMD	14.53 *(0.43)*	**4.19** *(0.86)*	11.74 *(0.44)*	7.25 *(0.69)*	12.69 *(0.47)*	6.53 *(0.78)*	**2.56** *(1.19)*
k-NN				2.93 *(0.83)*			

Table 8: C2+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	**2.54** *(0.86)*	2.59 *(0.86)*	2.72 *(0.85)*	2.59 *(0.86)*	**2.58** *(0.86)*	**2.58** *(0.87)*	2.61 *(0.86)*
WAD	3.34 *(0.75)*	2.59 *(0.87)*	2.62 *(0.84)*	2.77 *(0.80)*	2.77 *(0.79)*	2.66 *(0.85)*	3.66 *(0.75)*
WMD	6.75 *(0.66)*	3.84 *(0.90)*	16.77 *(0.42)*	5.34 *(1.02)*	6.27 *(0.84)*	4.26 *(1.33)*	**1.33** *(1.71)*
k-NN				2.27 *(1.00)*			

1. Analyzing the supervised classification procedures behaviors, DTM and WAD behave clearly better than WMD. However, WMD is the best method in presence of symmetric

Table 9: C3+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	4.26	4.11	4.14	4.11	4.22	4.10	4.18
	(0.72)	*(0.70)*	*(0.78)*	*(0.75)*	*(0.74)*	*(0.74)*	*(0.72)*
WAD	4.51	4.13	**3.97**	4.05	4.34	**3.94**	4.46
	(0.70)	*(0.72)*	*(0.77)*	*(0.75)*	*(0.74)*	*(0.75)*	*(0.72)*
WMD	12.75	5.04	14.88	7.25	11.17	6.53	**3.66**
	(0.42)	*(0.77)*	*(0.48)*	*(0.57)*	*(0.47)*	*(0.71)*	*(1.01)*
k-NN				3.17			
				(0.82)			

Table 10: C3+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	2.70	2.70	2.69	2.67	2.70	2.69	2.67
	(0.90)	*(0.89)*	*(0.90)*	*(0.90)*	*(0.90)*	*(0.90)*	*(0.89)*
WAD	3.31	**2.66**	2.69	2.72	2.88	**2.66**	3.23
	(0.85)	*(0.91)*	*(0.91)*	*(0.89)*	*(0.84)*	*(0.92)*	*(0.84)*
WMD	5.17	3.94	19.42	4.94	4.29	3.31	**1.65**
	(0.74)	*(0.85)*	*(0.47)*	*(0.78)*	*(0.89)*	*(1.36)*	*(1.41)*
k-NN				2.64			
				(0.92)			

Table 11: C4+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	1.86	1.73	1.78	1.78	1.90	**1.66**	1.82
	(1.00)	*(1.07)*	*(0.96)*	*(0.99)*	*(0.93)*	*(1.09)*	*(0.99)*
WAD	1.76	**1.57**	1.70	1.79	1.87	**1.57**	1.73
	(0.99)	*(1.13)*	*(1.03)*	*(0.96)*	*(0.95)*	*(1.09)*	*(0.99)*
WMD	10.21	4.18	12.08	3.95	8.40	3.39	3.47
	(0.46)	*(0.89)*	*(0.45)*	*(0.71)*	*(0.52)*	*(0.78)*	*(0.97)*
k-NN				0.69			
				(1.85)			

Table 12: C4+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	**0.34**	0.38	**0.34**	0.45	0.38	0.37	0.42
	(2.48)	*(2.36)*	*(2.36)*	*(2.03)*	*(2.26)*	*(2.43)*	*(2.23)*
WAD	0.50	0.40	**0.34**	0.43	0.51	0.38	0.50
	(1.82)	*(2.29)*	*(2.36)*	*(2.00)*	*(1.85)*	*(2.26)*	*(1.89)*
WMD	3.28	3.63	15.47	1.02	2.27	0.72	1.73
	(0.84)	*(1.00)*	*(0.39)*	*(1.58)*	*(1.04)*	*(1.59)*	*(1.59)*
k-NN				0.29			
				(2.60)			

Table 13: C5+DS1. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	2.59	2.56	2.29	2.19	2.30	2.18	2.64
	(1.37)	*(1.23)*	*(1.15)*	*(1.23)*	*(1.28)*	*(1.16)*	*(1.28)*
WAD	1.94	2.02	2.02	**1.89**	**1.84**	**1.76**	2.02
	(1.32)	*(1.30)*	*(1.21)*	*(1.28)*	*(1.32)*	*(1.34)*	*(1.26)*
WMD	24.40	3.39	12.99	23.57	20.59	6.75	2.19
	(0.26)	*(0.92)*	*(0.39)*	*(0.29)*	*(0.31)*	*(0.63)*	*(1.17)*
k-NN				0.99			
				(1.69)			

Table 14: C5+DS2. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	6.42	6.35	6.27	6.18	6.34	6.19	6.24
	(0.52)	*(0.53)*	*(0.54)*	*(0.50)*	*(0.52)*	*(0.54)*	*(0.52)*
WAD	6.16	6.16	6.40	6.10	**6.08**	**6.08**	**6.00**
	(0.53)	*(0.52)*	*(0.53)*	*(0.54)*	*(0.54)*	*(0.54)*	*(0.55)*
WMD	14.27	9.39	21.10	23.02	12.70	7.02	7.01
	(0.35)	*(0.52)*	*(0.28)*	*(0.29)*	*(0.38)*	*(0.59)*	*(0.58)*
k-NN				6.13			
				(0.53)			

magnitude contamination, both total and partial. This last result occurs when the depth function that WMD employs is KFSD.

2. Analyzing the behaviors of the depths and looking at the percentages in bold type, we observe that FSD, HMD and KFSD are the depths which are more involved in the best depth-based methods. However, due to the fact that we are considering different types of contamination, it is not clear which are the best depths, unlike the case of absence of contamination (see the considerations about Tables 1-4, point 3.), and they depend on the nature of the contamination.
3. This part of the simulation study shows a general robust behavior of k-NN, but also that in some cases the depth-based methods might be helpful. Actually, we would like to emphasize that a method involving a spatial depth, that is WMD+KFSD, performs very well in presence of the two cases of symmetric contaminations that we have con-

sidered, and that WAD+KFSD and WAD+FSD are among the best methods when there is shape contamination. On the other hand, the depth-based methods do not outperform k-NN in presence of asymmetric total and peaks magnitude contamination.

4. Finally, in presence of outliers the depth-based methods show the same pattern pointed out for the noncontaminated scenarios (see the considerations about Tables 1-4, point 4.), that is to say the differences in terms of performances with respect to k-NN shrink in presence of functional data with weak dependence structures.

5 REAL DATA STUDY: GROWTH DATA

To complete the comparison among the depth-based methods and k-NN, we also consider a real dataset. It consists in 93 growth curves: 54 correspond to heights of girls, 39 correspond to heights of boys (see Figure 4; for more details and other types of functional analyses about these data, see Ramsay and Silverman 2005).

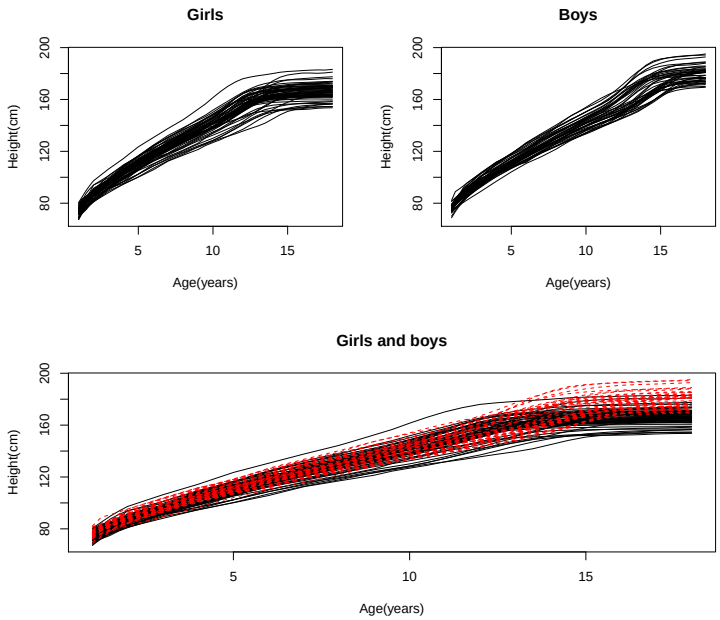

Figure 4: Growth curves: 54 heights of girls (top left), 39 heights of boys (top right), 93 heights of girls and boys (bottom; the dashed curves refer to boys).

In both cases, the curves are observed at a common discretized set of 31 nonequidistant ages between 1 and 18 years. We transform the initial dataset into an equally spaced and balanced dataset via linear interpolation, obtaining growth curves that are observed at a

common discretized set of 69 equidistant ages between 1 and 18 years. Clearly, other techniques can be used for this task, but we choose a simple linear interpolation because the main interest is classification, and not smoothing. Note that the same dataset has been used by López-Pintado and Romo (2006) and Cuevas et al. (2007) in their works about depth-based supervised functional classification. From our point of view, these data are interesting for two reasons: first, observing Figure 4, and in particular the heights of the girls, we can not discard the presence of some outlying curve; second, we work with the original data, in the sense that we do not use any registration technique to align the curves according to their velocity or acceleration peaks. This decision implies that we deal with all the variability of the original data, or in other words that we do not reduce the initial heterogeneity of the curves and the potential outlying behavior of some of them. For the previous reasons, it seems interesting to see how the spatial depths behave in this real situation. To perform the study, we consider 140 training samples composed by 40 and 30 randomly chosen curves of girls and boys, respectively. At each training sample we associate the test sample composed by the remaining 14 and 9 curves of girls and boys, respectively. For both methods and depths, we use the specifications described in Section 4. We report the performances of the 21 depth-based methods and of the k-NN procedure in the following table:

Table 15: Growth data. Means and coefficients of variation of the misclassification percentages.

Method	FMD	HMD	RTD	IDD	MBD	FSD	KFSD
DTM	14.81	10.06	20.00	19.41	17.11	19.13	12.55
	(0.60)	*(0.85)*	*(0.47)*	*(0.51)*	*(0.54)*	*(0.52)*	*(0.72)*
WAD	13.88	**8.76**	15.34	14.97	14.35	14.88	13.17
	(0.59)	*(0.89)*	*(0.55)*	*(0.56)*	*(0.57)*	*(0.56)*	*(0.61)*
WMD	29.57	**5.16**	14.10	31.18	26.18	17.76	**3.39**
	(0.39)	*(0.88)*	*(0.49)*	*(0.33)*	*(0.43)*	*(0.47)*	*(1.01)*
k-NN				3.88			
				(0.79)			

The best procedure in terms of mean misclassification percentage is given by WMD+KFSD (3.39%), which also was among the best pairs according to the simulation study results. There are other three procedures with percentages lower than 10%: the benchmark, k-NN (3.88%), and two methods involving the other kernel-based depth, i.e. WMD+HMD (5.16%) and WAD+HMD (8.76%). Therefore, with this real dataset, we observe that the kernel-based depths, especially together with WMD, are the unique depths that are able to compete with k-NN when the goal is to make supervised functional classification, and that

the best method involves KFSD.

6 CONCLUSIONS

In this article we have introduced two new functional depths: the functional spatial depth, FSD, and a kernelized functional spatial depth, KFSD. Both depths are based on a spatial approach, which represents an important original aspect with respect to the existing functional depths. The spatial approach was developed by Chaudhuri (1996) and Serfling (2002) in the multivariate context. We have proposed a functional extension of this approach which allows to study from a new point of view the depth of functional data.

The main novelty introduced by FSD consists in the connection between its definition and the notion of functional spatial quantiles. The theoretical study of the implications of this important relation should be part of our further research, as well as the study of the population and sample properties of FSD. On the other hand, with KFSD we have addressed the study of functional datasets that require analyses at a local level. KFSD depends on the choices about the kernel function, the kernel bandwidth and the distance function to consider, and therefore would require an accurate study in order to define procedures which take these decisions in a data-driven and efficient way. Thus, the natural next step in the development of KFSD should consist in defining these procedures.

FSD and KFSD have been used to solve supervised functional classification problems, especially in situations where the functional samples were contaminated since they contained some outlying curve, but we have also considered noncontaminated scenarios. We studied the classification performances of a benchmark procedure such as k-NN and of three depth-based methods, DTM, WAD and WMD. The three depth-based methods were used together with FSD, KFSD and five more existing functional depths. KFSD has proved to be the best depth if compared with the remaining ones, and a good competitor for k-NN, Moreover, other interesting results have been observed. First, k-NN was clearly the best procedure in absence of contamination but, in such nonfavourable situations for the depth-based methods, WAD+KFSD showed a stable and fairly satisfactory behavior. Second, WMD+KFSD was the best method in presence of symmetric-type contaminations, also if compared to the benchmark, which in its turn outperformed the depth-based methods in presence of

asymmetric-type contaminations. In presence of shape contamination, there was no a clear best procedure, but WAD+FSD and WAD+KFSD were among the best methods. Third, under noncontaminated and contaminated scenarios, we observed a shrinkage in the differences in terms of classification ability between k-NN and the depth-based methods. This shrinkage was in favour of the depth-based methods and occurred in presence of functional data with a weak dependence structure. Finally, the good KFSD performances were confirmed by the results of a real data study involving growth curves, which pointed out WMD+KFSD as the best classification method for this real dataset.

ACKNOWLEDGMENTS

This research was partially supported by Spanish Ministry of Education and Science grant 2007/04438/001, by Madrid Region grant 2011/00068/001, by Spanish Ministry of Science and Innovation grant 2012/00084/001 and by MCI grant MTM2008-03010.

References

Biau, G., Bunea, F., and Wegkamp, M. (2005), "Functional Classification in Hilbert Spaces," *IEEE Transactions on Information Theory*, 51, 2163–2172.

Brown, B. M. (1983), "Statistical Uses of the Spatial Median," *Journal of the Royal Statistical Society, Series B*, 45, 25–30.

Cardot, H., Cénac, P., and Zitt, P.-A. (2011), "Efficient and Fast Estimation of the Geometric Median in Hilbert Spaces With an Averaged Stochastic Gradient Algorithm." *Bernoulli*, in press.

Chaudhuri, P. (1996), "On a Geometric Notion of Quantiles for Multivariate Data," *Journal of the American Statistical Association*, 91, 862–872.

Chen, Y., Dang, X., Peng, H., and Bart, H. L. (2009), "Outlier Detection With the Kernelized Spatial Depth Function," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31, 288–305.

Cérou, F. and Guyader, A. (2006), "Nearest Neighbor Classification in Infinite Dimension," *ESAIM. Probability and Statistics*, 10, 340–355.

Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008), "The Random Tukey Depth." *Computational Statistics and Data Analysis*, 52, 4979–4988.

Cuevas, A., Febrero, M., and Fraiman, R. (2006), "On the Use of the Bootstrap for Estimating Functions With Functional Data," *Computational Statistics and Data Analysis*, 51, 1063–1074.

— (2007), "Robust Estimation and Classification for Functional Data via Projection-Based Depth Notions," *Computational Statistics*, 22, 481–496.

Cuevas, A. and Fraiman, R. (2009), "On Depth Measures and Dual Statistics. A Methodology for Dealing With General Data," *Journal of Multivariate Analysis*, 100, 753–766.

Epifanio, I. (2008), "Shape Descriptors for Classification of Functional Data," *Technometrics*, 50, 284–294.

Febrero, M., Galeano, P., and González-Manteiga, W. (2008), "Outlier Detection in Functional Data by Depth Measures, With Application to Identify Abnormal NOx Levels," *Environmetrics*, 19, 331–345.

Ferraty, F. and Vieu, P. (2003), "Curves Discrimination: a Nonparametric Functional Approach," *Computational Statistics and Data Analysis*, 44, 161–173.

— (2006), *Nonparametric Functional Data Analysis : Theory and Practice*, New York: Springer.

Fraiman, R. and Muniz, G. (2001), "Trimmed Means for Functional Data," *Test*, 10, 419–440.

Hall, P., Poskitt, D., and Presnell, B. (2001), "A Functional Data-Analytic Approach to Signal Discrimination," *Technometrics*, 43, 1–9.

Hastie, T., Buja, A., and Tibshirani, R. (1995), "Penalized Discriminant Analysis," *The Annals of Statistics*, 23, 73–102.

James, G. and Hastie, T. (2001), "Functional Linear Discriminant Analysis for Irregularly Sampled Curves," *Journal of the Royal Statistical Society, Series B*, 63, 533–550.

López-Pintado, S. and Romo, J. (2006), "Depth-Based Classification for Functional Data," in *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, eds. Liu, R., Serfling, R., and Souvaine, D. L., Providence: American Mathematical Society, DIMACS Series, pp. 103–120.

— (2009), "On the Concept of Depth for Functional Data," *Journal of the American Statistical Association*, 104, 718–734.

Marx, B. and Eilers, P. (1999), "Generalized Linear Regression on Sampled Signals and Curves: a P-Spline Approach," *Technometrics*, 41, 1–13.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer.

Serfling, R. (2002), "A Depth Function and a Scale Curve Based on Spatial Quantiles," in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, ed. Dodge, Y., Basel: Birkhaüser, pp. 25–38.

— (2006), "Depth Functions in Nonparametric Multivariate Inference," in *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, eds. Liu, R., Serfling, R., and Souvaine, D. L., Providence: American Mathematical Society, DIMACS Series, pp. 1–16.