working papers

UNIVERSIDAD CARLOS III DE MADRID

DEPTH-BASED CLASSIFICATION FOR FUNCTIONAL DATA

Sara López-Pintado and Juan Romo*

Abstract

Classification is an important task when data are curves. Recently, the notion of statistical depth has been extended to deal with functional observations. In this paper, we propose robust procedures based on the concept of depth to classify curves. These techniques are applied to a real data example. An extensive simulation study with contaminated models illustrates the good robustness properties of these depth-based classification methods.

Key Words: Functional data, classification, data depth.

*López-Pintado, Department of Statistics, Rutgers University, e-mail: saral@stat.rutgers.edu; Romo, Departamento de Estadística, Universidad Carlos III de Madrid, e-mail: juan.romo@uc3m.es.

Depth-based classification for functional data

Sara López-Pintado and Juan Romo

ABSTRACT. Classification is an important task when data are curves. Recently, the notion of statistical depth has been extended to deal with functional observations. In this paper, we propose robust procedures based on the concept of depth to classify curves. These techniques are applied to a real data example. An extensive simulation study with contaminated models illustrates the good robustness properties of these depth-based classification methods.

1. Introduction

Classification is an important problem when dealing with functional observations. Due to the huge technological advances and the growing complexity of experiments, curves are now the data frequently produced in different fields, such as biology, physics and economics. Classification of functions has been recently considered by several authors. For example, in ([**8**]) a penalized discriminant analysis is proposed. It is adequate for situations with many highly correlated predictors, as those obtained by discretizing a function. Nonparametric tools to classify a set of curves have been introduced in [**3**]. They calculate the posterior probability of belonging to a given class of functions by using a consistent kernel estimator. In addition, a new method for extending classical linear discriminant analysis to functional data has been analyzed in [**9**]. This technique is particularly useful when only fragments of the curves are observed. The problem of unsupervised classification or clustering of curves is addressed in [**10**], who elaborate a flexible model-based approach for clustering functional data. It is effective when the observations are sparse, irregularly spaced or occur at different time points for each subject. [**1**] consider unsupervised clustering of functions. They fit the functional data by $B-$splines and partition the estimated model coefficients using a $k-$means algorithm. In a related problem, [**7**] explore a functional data-analytic approach to perform signal discrimination. Many of these procedures are, however, highly sensitive to outliers. We propose in this paper two robust methods for supervised classification of curves based on the idea of depth for functional data.

A statistical depth expresses the "centrality"or "outlyingness"of an observation within a set of data (or with respect to a probability distribution) and provides a

criterion to order observations from center-outward. [**22**] and [**14**] introduced the earlier and most used definitions of depth in the multivariate context. Recently, new notions of depth for finite dimensional data have been proposed and all of them extensively studied (see, for example, [**13**], [**15**], [**16**], [**18**], [**21**], [**23**], [**24**] and [**25**]). Depth-based classification has been already analyzed in the literature for multivariate observations in, e.g., [**5**], [**6**], [**11**], [**12**] and [**20**]. Recently, the notion of depth has been extended to functional data. Fraiman and Muniz ([**2**]) have proposed a depth for functions defined as the integral of univariate depths. Alternatively, López-Pintado and Romo ([**17**]) have introduced a functional depth based on the proportion of bands including the curve graph.

Robustness is an interesting feature of the statistical methods based on depth. Robust techniques are even more useful for functional data than for multivariate observations because functional outliers can affect the statistical analysis in many different ways and they are not always easy to identify. The idea behind the depth-based classification methods proposed below is to construct a distance between an observation and the most representative data from each group. The functional depth measures how representative or interior is an observation within its group.

The rest of the article is organized as follows. In the next section we recall some ideas on depth for functional data. Section 3 presents two depth-based classification procedures for curves. A real data example is analyzed in section 4. The robustness of these new discrimination methods is illustrated in section 5 with an extensive simulation. Finally, section 6 collects the main conclusions of the paper.

2. Depth for functional data

The classification methods proposed below can be implemented with any depth defined for functional observations. We will use the idea of depth for functions introduced in [**17**]. It is a graph-based approach. Therefore, in this section we review some concepts about function graphs that will be used throughout the paper. Let $x_1(t), ..., x_n(t)$ be a set of real functions. For simplicity, we will assume that they belong to the space $C[0, 1]$ of continuous functions defined on the interval $[0, 1]$. The graph of a function x is the subset of $\mathbb{R}^2$

$$(2.1) \qquad\qquad G(x) = \{(t, x(t)) : t \in [0, 1]\}.$$

The band in $\mathbb{R}^2$ determined by k curves from the sample $x_1, ..., x_n$ is

$$
V(x_{i_1}, x_{i_2}, ..., x_{i_k}) = \left\{ (t, y) : \ t \in [0, 1], \ \min_{r=1,...,k} x_{i_r}(t) \le y \le \max_{r=1,...,k} x_{i_r}(t) \right\}
$$

$$
(2.2) \qquad = \left\{ (t, y) : \ t \in [0, 1] \ \ y = \alpha_t \min_{r=1,...,k} x_{i_r}(t) + (1 - \alpha_t) \max_{r=1,...,k} x_{i_r}(t), \right.
$$

$$
\left. \text{for some } \alpha_t \in [0, 1] \right\}.
$$

Figure 1a shows the band determined by two curves $V(x_1, x_2)$. In Figure 1b two additional curves, y_1 and y_2, are also represented. The red curve y_1 is included in the band, whereas the black one y_2 does not (since it is not always inside the band). The band can be determined by more than two curves. Thus, Figure 2 presents a band $V(x_1, x_2, x_3)$ defined by three curves.
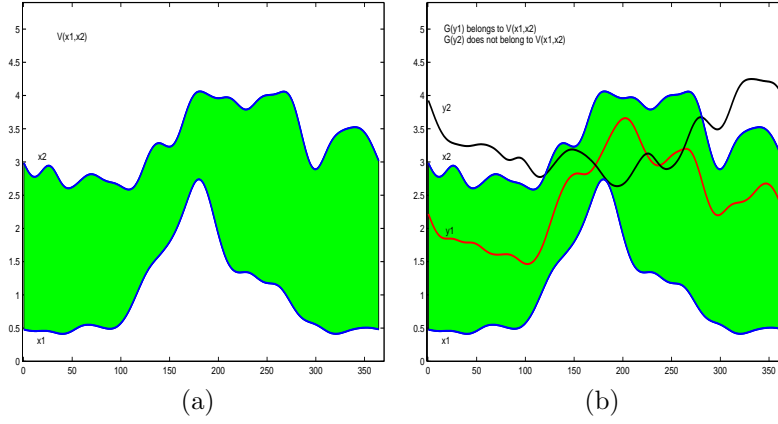
FIGURE 1. (a) Band defined by two curves x_1, x_2, and (b) a third curve y_1 belonging to the band $V(x_1, x_2)$ and a fourth curve y_2 not completely inside the band.
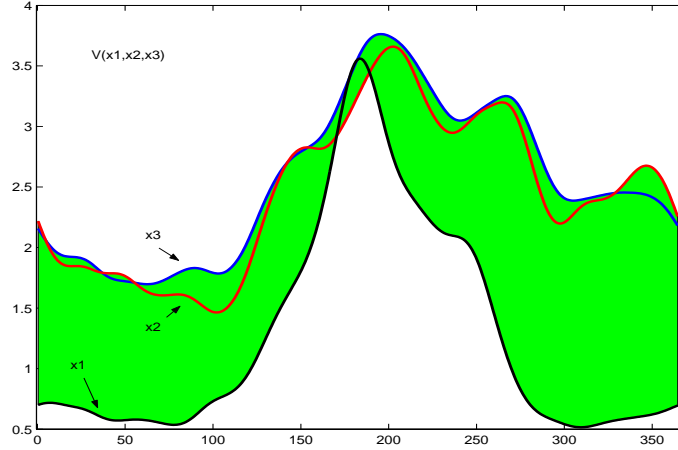

FIGURE 2. Band determined by three curves x_1, x_2, and x_3.

For any of the functions x in $\{x_1, ..., x_n\}$, the quantity

$$S_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \le i_1 < i_2 < ... < i_j \le n} I\{G(x) \subset V(x_{i_1}, x_{i_2}, ..., x_{i_j})\}, \ j \ge 2,$$

expresses the proportion of bands $V(x_{i_1}, x_{i_2}, ..., x_{i_j})$ determined by j different curves $x_{i_1}, x_{i_2}, ..., x_{i_j}$ containing the graph of x ($I\{A\} = 1$, if A occurs, and $I\{A\} = 0$, otherwise). The definition of depth for functional data introduced by [17] states that for functions $x_1, ..., x_n$, the *band depth* of any of these curves x is

$$S_{n,J}(x) = \sum_{j=2}^{J} S_n^{(j)}(x), \ J \ge 2.$$

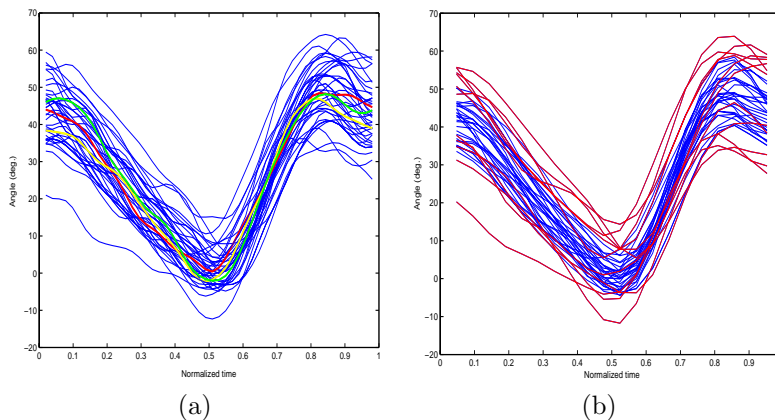(a)   (b)

FIGURE 3. (a) Three deepest curves in red, yellow and green, respectively, and (b) ten less deepest curves in red.

If $X_1, X_2, ..., X_n$ are independent copies of the stochastic process X which generates the observations $x_1, ..., x_n$, the population versions of these depth indexes are

$$S^{(j)}(x) = P\{G(x) \subset V(X_{i_1}, X_{i_2}, ..., X_{i_j})\}$$

and

$$S_J(x) = \sum_{j=2}^{J} S^{(j)}(x) = \sum_{j=2}^{J} P\{G(x) \subset V(X_{i_1}, X_{i_2}, ..., X_{i_j})\}.$$

A sample median function $\widehat{m}_{n,J}$ is a curve from the sample with highest depth value,

$$\widehat{m}_{n,J} = \underset{x \in \{x_1,...,x_n\}}{\arg\max} \; S_{n,J}(x),$$

and a population median is a function m_J in $C[0, 1]$ maximizing $S_J(\cdot)$. If there is more than one function achieving the depth maximum value, the median is defined as the average of those curves maximizing depth.

The band depth is applied to a a real data set of functions represented in Figure 3. The curves constitute the angle formed by the hip in the sagittal plane during one gait cycle for a set of thirty nine boys ([19]). The three deepest curves appear in red, yellow and green (left panel). Note that these curves resemble the main characteristics of the whole set. The ten less deepest curves are marked in red (right panel). These curves could be consider as outer observations within the sample.

Following also [17], a more flexible notion of depth (called *generalized band depth*) can be introduced if instead of considering the indicator function in the definition of depth we measure the set where the function is inside the corresponding band. For any function x in $\{x_1, ..., x_n\}$ and $j \geq 2$, let

$$A_j(x) \equiv A(x; x_{i_1}, x_{i_2}, ..., x_{i_j}) \equiv \left\{ t \in [0, 1] : \min_{r=i_1,...,i_j} x_r(t) \leq x(t) \leq \max_{r=i_1,...,i_j} x_r(t) \right\}$$
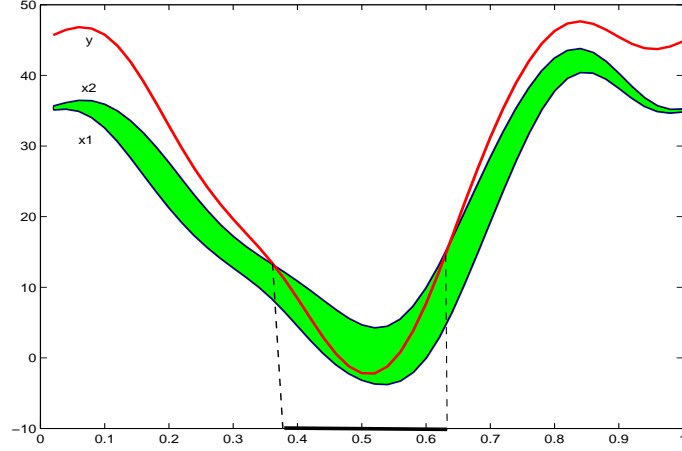
FIGURE 4. Set of points t where y is inside the band $V(x_1, x_2)$.

be the set of points in the interval $[0,1]$ where the function x is inside the band determined by the observations $x_{i_1}, x_{i_2}, ..., x_{i_j}$. Figure 4 shows a red curve y which is only partially contained in the band $V(x_1, x_2)$; the set where y is inside the band is marked in black.

If λ is the Lebesgue measure on the interval $[0,1]$, $\lambda(A_j(x))$ is the "proportion of time" that x is inside the band. Thus,

$$GS_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < i_2 < ... < i_j \leq n} \lambda(A(x; x_{i_1}, x_{i_2}, ..., x_{i_j})), \ j \geq 2,$$

is a generalized version of $S_n^{(j)}(x)$: if x is always inside the band, the measure $\lambda(A_j(x))$ is one and this generalizes the previous definition of depth. If a general compact interval I is used instead of $[0,1]$, the expression must be normalized by dividing by $\lambda(I)$.

As defined in [17], the *generalized band depth* of any of the curves x in $\{x_1, ..., x_n\}$ is

$$GS_{n,J}(x) = \sum_{j=2}^{J} GS_n^{(j)}(x), \ J \geq 2.$$

If $X_1, X_2, ..., X_n$ are independent copies of the process X which provide the observations $x_1, ..., x_n$, the population version of these indexes are

$$GS^{(j)}(x) = E \ \lambda(A(x; X_{i_1}, X_{i_2}, ..., X_{i_j})), \ j \geq 2,$$

and

$$GS_J(x) = \sum_{j=2}^{J} GS^{(j)}(x) = \sum_{j=2}^{J} E\lambda(A(x; X_{i_1}, X_{i_2}, ..., X_{i_j})), \ J \geq 2.$$

The generalized band depth is very stable in J, providing the same center-outward order in a collection of curves. Therefore, in what follows we consider $J = 2$ and denote $GS_2 = GS$. The theoretical properties of the band-depth and its generalized version are extensively studied in [17].
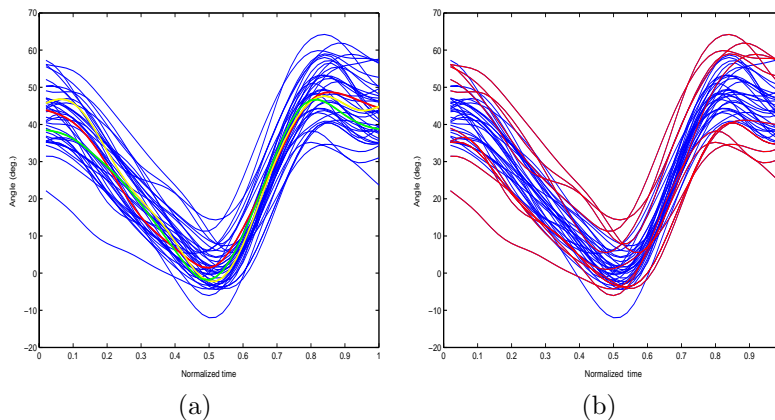
FIGURE 5. (a) Three deepest curves in red, yellow and green, and
(b) ten less deepest curves in red, using in both cases the general-
ized band depth.

The sample median function using the generalized band depth is denoted as

$$\widehat{m}_{n,GS} = \underset{x \in \{x_1, \ldots, x_n\}}{\arg\max} \; GS_{n,J}(x).$$

Figure 5 shows the deepest and less deepest curves for the generalized band
depth. For simplicity, we will denote by S any of these definitions of depth in the
rest of the paper.

In addition to the definition of median function, the notion of functional depth
allows to define order-based statistics, such as L-statistics. For example, a func-
tional version of the $\alpha-trimmed\ mean$ is defined as the average of the $n - [n\alpha]$
deepest curves from the sample ($[n\alpha]$ is the integer part of $n\alpha$). Let $x_{(1)}, \ldots, x_{(n)}$ be
the center-outward ordered sample, based on S, where $x_{(1)}$ is the deepest observa-
tion and $x_{(n)}$ is the less deepest one. The $\alpha-trimmed\ mean$ is defined as

$$m^{\alpha} = \frac{\sum_{i=1}^{n-[n\alpha]} x_{(i)}}{n - [n\alpha]}.$$

If there are curves in the sample with the same depth, $S(x_{i_1}) = S(x_{i_2}) = \ldots =
S(x_{i_k})$ with $i_1 < i_2 < \ldots < i_k$, they are ordered according to their original order
in the sample: if x_{i_1} is the $r - th$ deepest point from the sample ($x_{i_1} = x_{(r)}$) then
$x_{i_2} = x_{(r+1)}, \ldots, x_{i_k} = x_{(r+k-1)}$.

3. Classification for functions

In supervised classification, the curves belong to known groups $A_1, \ldots, A_G$ and
the goal is to assign any new observation to one of the G groups. Let $x_1, \ldots, x_n$
be a collection of continuous curves defined on the compact interval $[0, 1]$ and let
y_i be a natural number in $\{1, \ldots, G\}$ denoting the class of x_i. A classification
rule separates the space of continuous functions $C[0, 1]$ in G disjoint groups $\widehat{A}_1$,
$\widehat{A}_2, \ldots, \widehat{A}_G$, such that $x \in \widehat{A}_g$ if, and only if, the class prediction is g. Hence, a

classification rule assigns any new observation to a group based on the observations with known group.

To estimate the error (or misclassification) rate, the known observations are divided into the learning or training set $L = \{(x_1, y_1),\ (x_2, y_2), ...(x_{n_L}, y_{n_L})\}$ and the validation or test set $T = \{x_1, ..., x_{n_T}\}$. The classification rule is constructed from the training set and applied to the observations in the test set, comparing the true class membership with the predicted one. The estimated error rate e is the ratio between the number of misclassified observations in T and the total number n_T. In practice, there are many situations where the test set is not predefined and to estimate the error rate, random partitions of the sample data set into training L and test set T are considered.

In this section two new methods of supervised classification for functional data based on depth are presented: "distance to the trimmed mean" and "weighted averaged distance".

Distance to the trimmed mean (D). Calculates the distance from the new observation to the trimmed mean of each group. The steps are the following:

(1) Compute the trimmed mean for each group (for a given α): m_g^α, $g = 1, ..., G$.

(2) Calculate the distance from a new observation x_i to m_g^α: $d(x_i, m_g^\alpha)$, $g = 1, ..., G$, where d is any distance in $C[0, 1]$. In particular, throughout this paper we consider

$$d(x, y) = \|x - y\|_1 = \int_0^1 |x(t) - y(t)|\, dt$$

(3) Classify x_i in the group k such that: $d(x_i, m_k^\alpha) = \min_{g=1,...G} \left\{ d(x_i, m_g^\alpha) \right\}.$

Depending on the trimming, the two extreme cases are: i) to compute the distance to the mean ($\alpha = 0$), and ii) to calculate the distance to the deepest observation ($\alpha = (n - 1)/n$).

Weighted Averaged Distance (AD). This second classification method obtains the distance from an observation to a group as a weighted average of distances to each element in the group. The weights will be determined by the points depth within the group. Thus, the influence of any observation on the final distance will depend on its depth, i.e., the deeper an observation, the greater its weight in the calculated distance. Let $A_g = \left\{ x_1, ..., x_{n_g} \right\}$. The weighted average distance of x to A_g is

$$(3.1) \qquad AD(x, A_g) = \frac{\sum_{i=1}^{n_g} d(x, x_i) S(x_i)}{\sum_{i=1}^{n_g} S(x_i)}, \quad \text{where } n_g \text{ is the size of the group } A_g.$$

A problem with the notion of "weighted averaged distance" is that the result depends strongly on the number of observations in each group. Therefore, if the number of observations in separate groups are too different, the outcome of the classification is inaccurate. A plausible approach, denoted as "Trimmed Weighted Averaged Distance" (TAD), is to fix $m \leq n_1, ..., n_g$ and consider only the m deepest observations from each group to compute the distance. Thus,
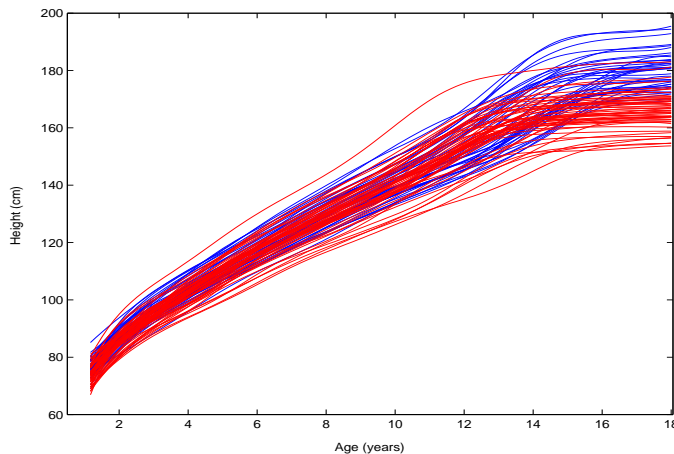
FIGURE 6. Growth curves for fifty four girls (in red) and thirty nine boys (in blue).

$$(3.2) \qquad TAD(x, A_g) = \frac{\sum\limits_{i=1}^{m} d(x, x_{(i)}) S(x_{(i)})}{\sum\limits_{i=1}^{m} S(x_{(i)})}.$$

4. Real data example

Next we apply the classification techniques to the growth curves for a sample of boys and girls from [**19**] (see Figure 6). The observations are the heights (in centimeters) of fifty four girls and thirty nine boys measured at a set of twenty nine ages from one to eighteen years old. The original data was smoothed using a spline basis. We use this data set to compare the performance of the classification methods introduced in the previous section (D, AD and TAD), applying them to classification in two groups (boys-girls). We have estimated the error or misclassification rate by separating the sample into k groups of the same size and considering one of them as the validation or test group T. The remaining observations constitute the learning set L. We calculate the number of misclassified observations from the validation group. All these steps are repeated changing the test group k times and the total error rate is defined as the number of misclassified observations over the total number of observations for each method. Three validation approaches are used to estimate the probability error: V1 (dividing the sample in $k = 4$ parts), V2 (dividing the sample in $k = 10$ parts) and V3 (classical cross-validation, dividing the sample in $k = n_1 + n_2$ parts, where n_i is the size of group i).

a. Distance to the trimmed mean (D)

The $\alpha-$trimmed mean is computed for each group of curves (boys and girls respectively) with $\alpha = 0.2$ and using three different notions of depth: band depth with $J = 3$ (S_3), band depth with $J = 4$ (S_4) and the generalized band depth (GS). The error rate is estimated using three validation procedures V1, V2 and

TABLE 1. Error rate for the distance to the trimmed mean method D.

	DS3	DS4	DGS	DM
V1	0.2374	0.2374	**0.1486**	0.2260
V2	0.2488	0.2113	**0.1768**	0.2440
V3	0.1828	0.1828	**0.1613**	0.2258

TABLE 2. Error rate for the weighted averaged distance method AD.

	ADS3	ADS4	ADGS	DM
V1	0.2622	0.2508	**0.1961**	0.2395
V2	0.2565	0.2565	**0.1893**	0.2315
V3	0.2473	0.2473	**0.1935**	0.2258

TABLE 3. Error rate for the trimmed weighted averaged distance method TAD.

	TADS3	TADS4	TADGS	DM
V1	0.2601	0.2601	**0.2075**	0.2260
V2	0.2536	0.2536	**0.1690**	0.2363
V3	0.2436	0.2436	**0.1690**	0.2258

V3. Table 1 shows the error rates comparing the distance to the $\alpha-$trimmed mean (three first columns) and the distance to the mean.

b. Weighted averaged distance (AD)

Compute each data depth within its group and obtain the weighted averaged distance of every observation to each group as in (3.2). In a similar way to part a, we consider three validation procedures V1, V2 and V3 and the previous three definitions of depth (Table 2).

c. Trimmed weighted averaged distance (TAD)

This third classification method is a slight modification of AD. We consider only the $m = 39$ deepest curves from each group to calculate the weighted averaged distance as in (3.2). The results of this discrimination procedure appear in Table 3.

The error rates are minimized for the generalized band depth in all three classification procedures (D, AD and TAD). Moreover, the best error rate is obtained with procedure D and depth GS.

In addition to calculating the error rate for the boys-girls growth curves example, we have also estimated the error distribution. One third of the sample (growth curves for boys and girls together) is randomly used as the test group and the remaining observations constitute the learning or training set. We have repeated this $r = 50$ times and the error rate is computed each time for the different classification methods. In this example is convenient to apply TAD instead of AD because the sizes of the two groups of curves are not similar (fifty four girls versus thirty nine boys). The box-plots of the error distributions are represented in Figure 7 and the mean, first, second and third quartiles appear in Table 4.
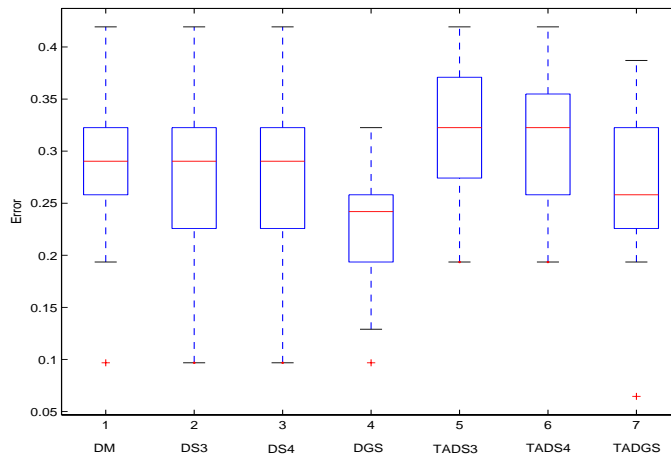
FIGURE 7. Error rate distribution for boys and girls growth curves. The classification procedures are: distance to the mean (DM), distance to the 0.2-trimmed mean with S_3, S_4, GS, (DS_3, DS_4, DGS) and trimmed weighted averaged distance with S_3, S_4, GS, ($TADS3$, $TADS4$ and $TADGS$).

TABLE 4. Mean, first, second and third quartiles of the error rate distribution for the growth data.

	mean	q1	q2	q3
DM	0.2839	0.2581	0.2903	0.3226
DS3	0.2710	0.2258	0.2903	0.3226
DS4	0.2694	0.2258	0.2903	0.3226
DGS	**0.2274**	**0.1935**	**0.2419**	**0.2581**
TADS3	0.3177	0.2742	0.3226	0.3710
TADS4	0.3097	0.2581	0.3226	0.3548
TADGS	0.2645	0.2258	0.2581	0.3226

The best results (minimum error rates) correspond to the distance to the trimmed mean, followed by the trimmed weighted averaged distance based both of them on the generalized band-depth.

5. Simulation Results

We now illustrate the robustness of the classification techniques for functional data introduced above. To compare their behavior, we have simulated curves from several models with some type of contamination. There are different ways of contaminating a continuous process; among others, we will use the ones in [**2**]. In all cases analyzed below we have considered two models and simulated seventy curves from each one. The training set consists of thirty randomly chosen curves from each group and the remaining forty observations from each sample are the test group. This is repeated $r = 50$ times and the error rates are computed. The methods used to classify are DM (distance to the mean), DS3 (distance to the trimmed mean based on S_3), DS4 (distance to the trimmed mean based on S_4), DGS (distance
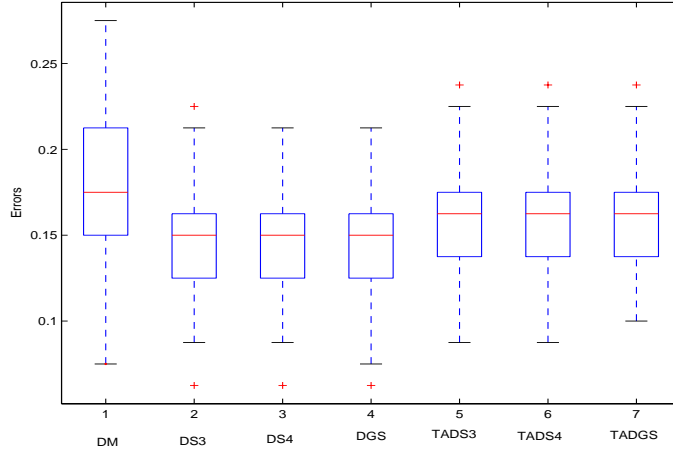
FIGURE 8. Box-plots of the error distribution for curves simulated from model A with $\alpha = 0.2$, $q = 0.05$ and $M = 25$.

to the trimmed mean based on GS), TADS3 (trimmed averaged distance based on S_3), TADS4 (trimmed averaged distance based on S_4) and TADGS (trimmed averaged distance based on GS). In all cases, the trimmed averaged distance is calculated with $m = 30$.

Model A (Asymmetric contamination). Consider two gaussian processes with the same covariance function but different mean function, and with some contaminated observations in the second process,

$$
\begin{aligned}
X_i(t) &= f(t) + \varepsilon_i \\
Y_i(t) &= g(t) + \varepsilon_i + c_i M,
\end{aligned}
$$

where ε is a gaussian process with zero mean and covariance function

(5.1) $$\gamma(t, s) = 0.25 * \exp\left\{ -|t - s|^2 \right\},$$

$f(t) = 4t$, $g(t) = 7t$, c_i takes the value 1 with probability q and 0 with probability $1 - q$ and M is the contamination size. The simulation results are obtained following the procedure described above. Figure 8 and Table 5 show the errors distributions based on the different classification methods introduced previously. The trimmed mean is computed with $\alpha = 0.2$, $q = 0.1$ and the contamination size M is 25. The best error rate is obtained with DGS followed by TADGS.

Model A simulation has been repeated with $q = 0.1$ and the error distribution is represented in Figure 9. The best results again correspond to DGS and $TADGS$.

Model B (Symmetric contamination). A symmetric contamination can be introduced by considering

$$
\begin{aligned}
X_i(t) &= f(t) + \varepsilon_i \\
Y_i(t) &= g(t) + \varepsilon_i + c_i \sigma_i M,
\end{aligned}
$$

where ε_i is a gaussian process with covariance function (5.1), c_i and M are defined as in model A and σ_i is a sequence of random variables independent from c_i taking

TABLE 5. Mean, first, second and third quartiles of the error distribution for model A with $q = 0.05$.

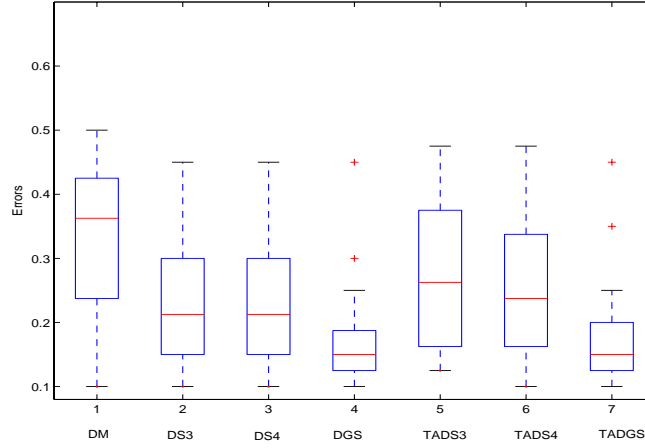|	mean	q1	q2	q3
DM	0.1772	0.15	0.175	0.2125
DS3	0.1455	0.125	0.15	0.1625
DS4	0.1458	0.125	0.15	0.1625
DGS	**0.1450**	**0.125**	**0.15**	**0.1625**
TADS3	0.16	0.1375	0.1625	0.1750
TADS4	0.1597	0.1375	0.1625	0.1750
TADGS	0.1605	0.1375	0.1625	0.1750

FIGURE 9. Box-plots of the error distribution for curves simulated from model A with $\alpha = 0.2$, $q = 0.1$ and $M = 25$.

TABLE 6. Mean, first, second and third quartiles of the error distribution in model B.

	mean	q1	q2	q3
DM	0.1970	0.1167	0.1667	0.2417
DS3	0.1567	0.1333	0.15	0.1833
DS4	0.1528	0.1250	0.15	0.1750
DGS	**0.1447**	**0.1167**	**0.15**	**0.1667**
TADS3	0.1677	0.1333	0.1667	0.2
TADS4	0.1630	0.1333	0.1667	0.1917
TADGS	0.1572	0.1333	0.1583	0.1833

values 1 and -1 with probability $1/2$. Figure 10 and Table 6 show the error distribution using different classification procedures (DM, DS_3, DS_4, DGS, $TADS_3$, $TADGS$) with $\alpha = 0.2$, $q = 0.1$ and $M = 25$. The minimum error rate is obtained with DGS followed by DS3.
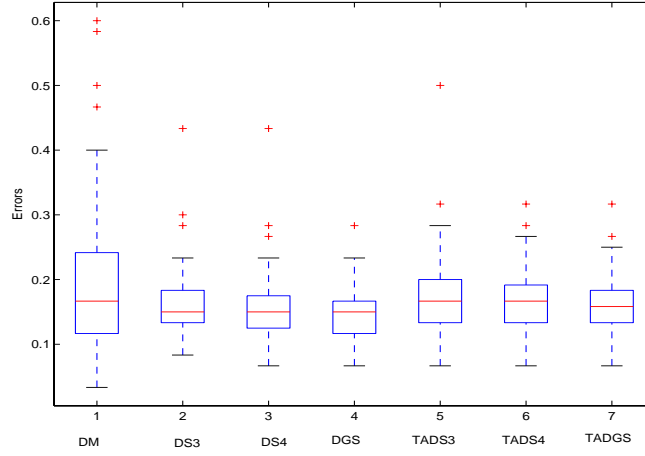
FIGURE 10. Box-plots of the error distribution for data simulated from model B with $\alpha = 0.2$, $q = 0.1$, and $M = 25$.
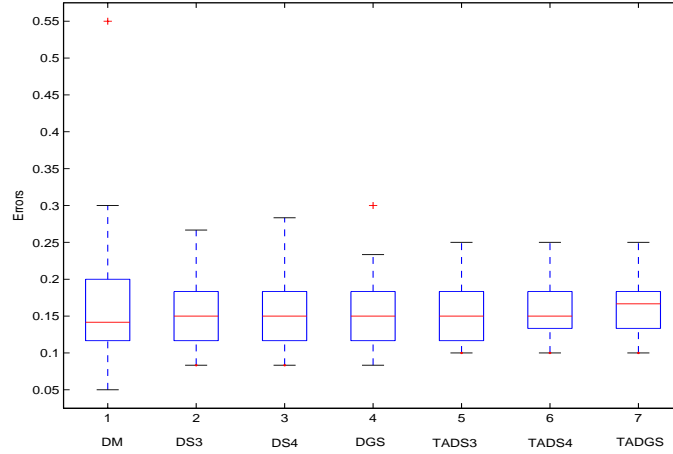

FIGURE 11. Box-plots of the error distribution for simulated data from model C with $\alpha = 0.2$, $q = 0.1$, and $M = 25$.

Model C (Partially contaminated model). The partially contaminated model can be expressed as

$$
\begin{aligned}
X_i(t) &= f(t) + \varepsilon_i \\
Y_i(t) &= \begin{cases} g(t) + \varepsilon_i + c_i \sigma_i M & \text{for } t \geq T_i \\ g(t) + \varepsilon_i, & \text{for } t < T_i, \end{cases}
\end{aligned}
$$

where c_i, σ_i and M are defined as above and T_i is a random number generated from a uniform distribution on $[0, 1]$.

The box-plots of the error distribution for data simulated from model C can be seen in Figure 11. The mean, first, second and third quartiles of the error distribution are represented in Table 7. In this case, the best result is obtained with DS3.

Table 7. Mean, first, second and third quartiles of the error distribution for model C.

	mean	q1	q2	q3
DM	0.1617	0.1167	**0.1417**	0.2
DS3	**0.1540**	**0.1167**	0.15	**0.1833**
DS4	0.1563	0.1167	0.15	0.1833
DGS	0.1577	0.1167	0.15	0.1833
TADS3	0.1567	0.1167	0.15	0.1833
TADS4	0.1573	0.1333	0.15	0.1833
TADGS	0.1663	0.1333	0.1667	0.1833
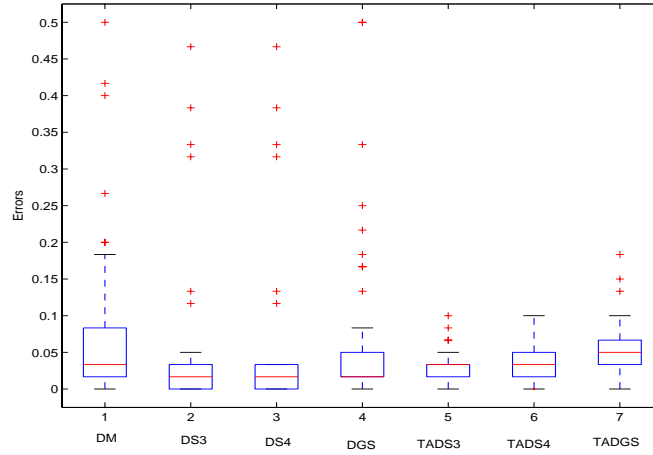
Figure 12. Box-plots of the error distribution for simulated data from model D with $\alpha = 0.2$, $q = 0.1$, and $M = 25$.

Model D (Peaks contaminated model). The fourth model is contaminated with peaks and is given by

$$
\begin{aligned}
X_i(t) &= f(t) + \varepsilon_i \\
Y_i(t) &= g(t) + \varepsilon_i + c_i M \qquad \text{for} \ \ T_i \le t \le T_i + l \quad \text{and} \\
Y_i(t) &= g(t) + \varepsilon_i \qquad\qquad \text{for} \ t \notin [T_i, T_i + l],
\end{aligned}
$$

where c_i is defined as in previous model, $l = 2/30$ and T_i is a random number generated from a uniform distribution on $(0, 1 - l)$. The idea underlying this model is to contaminate it only during a short interval of length l.

As shown in Figure 12 and Table 8, in model D the methods based on S_3 and S_4 give, in contrast with models A-C, significantly better error rates than the ones based on GS. This is caused by the type of contamination considered in this last case; when the contamination occurs only in a short interval of the domain of the curve, the band depth of the contaminated curves is low, but the generalized band depth can still be high by definition. Therefore, classification procedures based on GS are less robust in this situation.

TABLE 8. Mean, first, second and third quartiles of the error distribution for model D.

	mean	q1	q2	q3
DM	0.0710	0.0167	0.0333	0.0833
DS3	0.0487	0	0.0167	0.0333
DS4	0.0477	**0**	**0.0167**	**0.0333**
DGS	0.077	0.0167	0.0167	0.05
TADS3	**0.0303**	0.0167	0.0333	0.0333
TADS4	0.0310	0.0167	0.0333	0.05
TADGS	0.0540	0.0333	0.05	0.0667

Interestingly, in all models considered (A-D), the depth-based classification methods perform better than the distance to the mean (which is the worst method due to contamination).

6. Conclusions

Nonparametric robust classification procedures for functional data have been proposed and analyzed. These methods are based on the notion of depth for functional observations and the definitions introduced in [**17**] have been used throughout the paper. The idea of depth provides a criterion to order a sample of curves from center-outward and robust location estimates, such as the trimmed mean, are defined for functional data. Classic discriminant techniques (for example calculating the distance to the mean curve in each group) can be modified by using the trimmed mean instead of the mean as a representative element from each cluster. In addition, a second method of classification using a weighted averaged distance, with weights depending on depth, is also proposed. The good behavior in terms of robustness of the new procedures is illustrated using a real data example and also simulated data from contaminated processes. Interesting topics that could be addressed in the future are the analysis of some theoretical properties of these new classification procedures and their comparison with those already studied in the literature (see e.g. [**3**], [**4**] and [**9**]). Also, depth-based clustering methods for functional data extending the ideas in [**10**] and [**11**] could be considered.

References

[1] C. Abraham, P. A. Cornillon, E. Matzner and N. Molinari, Unsupervised curve clustering using B-splines, *Scandinavian Journal of Statistics* **30** (2003), 581-596.

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

[3] F. Ferraty and P. Vieu, Curves discrimination: a nonparametric functional approach, *Computational Statistics and Data Analysis* **44** (2003), 161–173.

[4] F. Ferraty and P. Vieu, Nonparametric models for functional data with applications in regression, time series prediction and curves discrimination, *Journal of Nonparametric Statistic* **16** (2004), 111–125.

[5] A. K. Ghosh and P. Chaudhuri, On maximum depth and related classifiers, *Scandinavian Journal of Statistics* **32** (2005), 327–350.

[6] A. K. Ghosh and P. Chaudhuri, On data depth and distribution-free discriminant analysis using separating surfaces, *Bernoulli* **11** (2005), 1–27.

[7] P. Hall, D. S. Poskitt and B. Presnell, A functional data-analytic approach to signal discrimination, *Technometrics* **43** (2001), 1–9.

[8] T. Hastie, A. Buja and R. J. Tibshirani, Penalized discriminant analysis, *Annals of Statistics* **23** (1995), 73–102.

[9] G. James and T. Hastie, Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society*, Series B **63** (2001), 533–550.

[10] G. James and C. Sugar, Clustering for sparsely sampled functional data. *Journal of the American Statistical Association* **98** (2003), 397–408.

[11] R. Jornsten, Clustering and classification based on L_1 data depth. *Journal of Multivariate Analysis* **90** (2004), 67–89.

[12] R. Jornsten, Y. Vardi and C. H. Zhang, A robust clustering method and visualization tool based on data depth. In Y. Dodge editor, *Statistical Data Analysis*, pp. 353–366. Basel: Birkhauser, 2002.

[13] G. Koshevoy and K. Mosler, Zonoid trimming for multivariate distributions. *Annals of Statistics* **25** (1997), 1998–2017.

[14] R. Liu, On a notion of simplicial depth. *Proceedings National Academy U.S.A.* **85** (1988), 1732-1734.

[15] R. Liu, On a notion of data depth based on random simplices. *Annals of Statistics* **18** (1990), 405–414.

[16] R. Liu, J. M. Parelius and K. Singh, Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics* **27** (1999), 783–858.

[17] S. López-Pintado and J. Romo, On the concept of depth for functional data. Preprint, 2005.

[18] H. Oja, Descriptive statistics for multivariate distributions. *Statistics and Probability Letters* **1** (1983), 327–332.

[19] J. O. Ramsay and B. W. Silverman *Functional Data Analysis.* Springer-Verlag, New York, 1997.

[20] I. Ruts and P. J. Rousseeuw, Computing depth contours of bivariate point clouds. *Computational Statistics and Data Analysis* **23** (1996), 153–168.

[21] K. Singh, A notion of majority depth. Unpublished manuscript.

[22] J. W. Tukey, Mathematics and picturing of data. *Proceedings of the 1974 International Congress of Mathematicians, Vancouver.* (1975) **2**, 523–531.

[23] Y. Vardi and C. H. Zhang, The multivariate L_1-median and associated data depth. *Proceedings of the National Academy of Science, USA* **97** (2000), 1423–1426.

[24] Y. Zuo, Projection based depth functions and associated medians. *Annals of Statistics* **31** (2003), 1460–1490.

[25] Y. Zuo and R. Serfling, General notions of statistical depth function. *Annals of Statistics* **28** (2000), 461–482.

DEPARTMENT OF STATISTICS, UNIVERSIDAD CARLOS III DE MADRID, 28903 GETAFE, MADRID, SPAIN
    E-mail address: saral@est-econ.uc3m.es

DEPARTMENT OF STATISTICS, UNIVERSIDAD CARLOS III DE MADRID, 28903 GETAFE, MADRID, SPAIN
    E-mail address: juan.romo@uc3m.es