DEPARTMENT OF DECISION SCIENCES AND INFORMATION MANAGEMENT (KBI)

Multivariate Functional Halfspace Depth

Gerda Claeskens, Mia Hubert, Leen Slaets and Kaveh Vakili [1]

December 17, 2012

A multivariate depth for functional data is defined and studied. By the multivariate nature and by including a weight function, it acknowledges important characteristics of functional data, namely differences in the amount of local amplitude, shape and phase variation. Both population and finite sample versions are studied. The multivariate sample of curves may include warping functions, derivatives and integrals of the original curves for a better overall representation of the functional data via the depth. A simulation study and data example confirm the good performance of this depth function.

Keywords: statistical depth, functional data, time warping, multivariate data.

1 Introduction

Nowadays, functional data are frequently observed and many statistical methods have been developed to retrieve useful information from these data sets. Typically, the observed data consist of a set of N curves, each measured at different time points

$t_1, \ldots, t_T$. For an overview, see Ramsay and Silverman (2006); Ferraty and Vieu (2006). Basic questions of interest in functional data analysis (FDA) are (i) the estimation of the central tendency of the curves, (ii) the estimation of the variability among the curves, (iii) the detection of outlying curves, as well as (iv) classification or clustering of such curves.

In this paper we consider *multivariate functional data*. We observe for all observation units at each time point a K-dimensional vector of measurements, which arise from an underlying set of K curves. A popular example is the bivariate gait data set, which contains the simultaneous variation of the hip and knee angles for 39 children at 20 equally space time points (Ramsay and Silverman, 2006). Berrendero et al. (2011) have $K = 3$ when recording daily temperature functions at 3, 9 and 12cm below the surface during $N = 21$ days. Sangalli et al. (2009) and Pigoli and Sangalli (2012) present several multivariate functional data from medical studies.

Different types of multivariate functional data arise by computing additional curves, starting from one observed set of univariate functional data. A well-studied situation is the addition of the first order *derivatives* which provides additional information on the shape of the curves and consequently is interesting to detect curves with an outlying shape (Cuevas et al., 2007). Note that this is different from a common practice in chemometrics, where observed spectral data are often *replaced* by their first-order derivatives in order to eliminate baseline features. Also higher order derivatives could be added. This has been applied thoroughly in the Berkeley growth data set (Ramsay and Silverman, 2006), which contains the heights of children and the estimated acceleration curves that correspond to the second-order derivatives.

In this paper we introduce to the depth calculation the inclusion of other functions of the original set of curves (such as warping functions, derivatives, integrals,...) which allows us to obtain more powerful conclusions about the data-driven process.

Some interesting functions are obtained from a *warping* procedure, which often precedes the analysis of functional data. Typically some warping method (also known as curve alignment) is applied to the observed curves as a preprocessing step, but no further information is retained from this analysis. In Slaets et al. (2012) it is shown how the information from the warping procedure can be incorporated into a clustering procedure of functional data. In Section 4.3 we show the benefits of a multivariate analysis of the warped data together with the curves obtained via the warping function.

A different augmentation of the data is presented in Section 3. It contains the analysis of a real data set which consists of acceleration signals over time from an industrial machine (De Ketelaere et al., 2011). Most of the observed curves, see Figure 1(a), follow a similar nonlinear pattern, but we also notice several curves with a deviating trend, most prominently at the final stage of the production. Additionally to these acceleration signals, we do not use their derivatives but rather the *integrated* curves as they represent the underlying velocity, see Figure 1(b). Also here, we see a global structure as well as deviating signals. On both plots we have added the cross-sectional mean curve, colored orange. Further we have plotted our new estimator for the central tendency of the curves, shown in dark red color. It is already obvious that these estimates are less influenced by the outlying curves. For the velocity curves, the effect is less pronounced as the outlying curves occur in both directions of the central pattern.

Our approach to estimate the central tendency of multivariate functional data is based on the concept of depth. Depth functions were initially defined for multivariate data. They provide an ordering from the center outwards such that the most central object gets the highest depth value and the least central objects the smallest depth. More recently, several notions of depth have been proposed for univariate functional
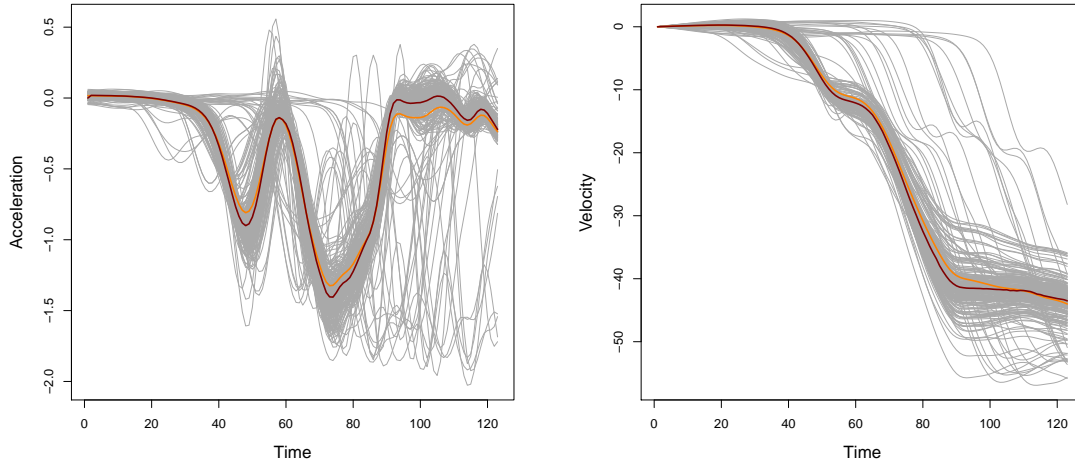
3

Figure 1: (a) Acceleration and (b) velocity signals, with cross-sectional mean curve (in orange) and depth-based median curve (in dark red).

data, such as the Fraiman and Muniz depth (FM, Fraiman and Muniz, 2001), the h-mode and random projection depth (RP, Cuevas et al., 2007), the band depth and modified band depth (MBD, López-Pintado and Romo, 2009) and the half-region depth (López-Pintado and Romo, 2011). The FM depth and MBD depth are quite similar, as they both consider a (univariate) depth function at each time point t and define the functional depth as the *average* of these depth values over all time points. Cuevas et al. (2007) have proposed to consider the curves and their derivatives, yielding the bivariate random projection depth (RPD). For a number of random projections, they project both sets of curves on each direction, apply a multivariate depth function on the bivariate sample and finally average the depth values over the random projections.

We generalize several of these ideas by constructing a depth function for K-variate curve samples, which we define the multivariate functional depth (MFD). Our definition also averages a multivariate depth function over the time points, but in addition

it includes a weight function, which accounts for variability in amplitude, to adapt to the functional nature of the data. More specifically we choose Tukey's halfspace depth (Tukey, 1975) as the building block, which leads to the multivariate functional halfspace depth (MFHD).

The population and finite-sample definition of MFD and MFHD and their main properties are given in Section 2. We define and characterize the MFHD median, as the curve with maximal MFHD. In Section 3 we illustrate on the industrial data set how this new depth concept can be used to estimate the central tendency of the curves, as well as the variability among the curves. In Section 4 we provide the results of a simulation study in which we compare several depth functions and several augmented data sets (such as derivatives and warping functions). Section 5 concludes and gives directions for further research. All proofs are collected in the Appendix.

2 Definition and properties of multivariate functional depth

2.1 Notation

Consider a K-variate (finite K), real-valued stochastic process of continuous functions $\boldsymbol{\mathcal{Y}} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_K)$ with for $j = 1, \ldots, K$, $\mathcal{Y}_j : U \to \mathbb{R} : t \mapsto \mathcal{Y}_j(t)$ continuous on a compact interval U and denote its cumulative distribution by $F_{\boldsymbol{\mathcal{Y}}}$. Thus, for every finite set of time points $t_1, \ldots, t_T \in U$, $(\boldsymbol{\mathcal{Y}}(t_1), \ldots, \boldsymbol{\mathcal{Y}}(t_T))$ is a random variable on $\left(\mathbb{R}^K\right)^T$ and at each time point $t \in U$, $\boldsymbol{\mathcal{Y}}(t)$ is a K-variate random variable with associated cumulative distribution function (cdf) $F_{\boldsymbol{\mathcal{Y}}(t)}$.

Real numbers, vectors, continuous functions on an interval U and vectors of functions are all used in conjunction with each other. To avoid confusion, we provide an

overview of the notation that is used throughout this paper. The set of continuous functions on U is denoted by $\mathcal{C}(U)$. Elements thereof and their graphs are denoted by capital letters (e.g. X). For K-vectors of continuous functions in $\mathcal{C}(U)^K$ and their graphs, bold capital letters are used (e.g. $\boldsymbol{X}$) or the vector notation $(X_1, X_2, \ldots, X_K)$ where $X_i \in \mathcal{C}(U)$. The function value of a curve X at a time point t is denoted by $X(t) \in \mathbb{R}$. The vector of function values of an element $\boldsymbol{X}$ in $\mathcal{C}(U)^K$ at a time point t is denoted by $\boldsymbol{X}(t) = (X_1(t), \ldots, X_K(t))$. The empirical cumulative distribution function based on a sample $\{\boldsymbol{Y}_1(t), \ldots, \boldsymbol{Y}_N(t)\}$ each with the same distribution as $\boldsymbol{\mathcal{Y}}(t)$ is denoted by $F_{\boldsymbol{\mathcal{Y}}(t),N}$. For vectors in $\mathbb{R}^K$, bold lowercase letters are used (e.g $\boldsymbol{a}$) or the vector notation $(a_1, a_2, \ldots, a_K) \in \mathbb{R}^K$. For matrices capital letters early in the alphabet are used (e.g $\boldsymbol{A}$), while later letters (e.g. X) are reserved for curves. For real numbers, lowercase letters are used (e.g. a).

2.2 Population definition

2.2.1 A general multivariate depth as building block

A depth function provides an ordering from the center outwards such that the most central objects get the highest depth and the least central objects the smallest depth. Let $D(\cdot; F_{\mathcal{X}}) : \mathbb{R}^K \to [0, 1]$ be a statistical depth function for the probability distribution of a K-variate random vector $\mathcal{X}$ with cdf $F_{\mathcal{X}}$, according to Zuo and Serfling (2000a). Associated with the depth function is the depth region $D_\alpha(F_{\mathcal{X}})$ at level $\alpha \geqslant 0$, defined as $D_\alpha(F_{\mathcal{X}}) = \{\boldsymbol{x} \in \mathbb{R}^K : D(\boldsymbol{x}; F_{\mathcal{X}}) \geqslant \alpha\}$.

The new definition of multivariate *functional* depth combines the local depths of $\boldsymbol{\mathcal{Y}}(t)$ at each time point $t \in U$ and takes the local changes in the amount of variability in amplitude (vertical variability) into account by including a weight function.

Definition 1. *Consider a K-variate stochastic process $\{\boldsymbol{\mathcal{Y}}(t), t \in U\}$ on $\mathbb{R}^K$ with cdf*

$F_{\boldsymbol{\mathcal{Y}}}$ that generates continuous paths in $\mathcal{C}(U)^K$. Let D be a statistical depth function on $\mathbb{R}^K$ and $\alpha \in (0,1]$ such that $\int_U \mathrm{vol}\{D_\alpha(F_{\boldsymbol{\mathcal{Y}}(u)})\}du \in (0, +\infty)$. Take an arbitrary $\boldsymbol{X} \in \mathcal{C}(U)^K$. The multivariate functional depth (MFD) of $\boldsymbol{X}$ is defined as

$$MFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) \;\; = \;\; \int_U D(\boldsymbol{X}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) \cdot w_\alpha(t; F_{\boldsymbol{\mathcal{Y}}(t)}) \, dt, \qquad (1)$$

for a fixed choice of $\alpha \in (0,1]$ with

$$w_\alpha(t; F_{\boldsymbol{\mathcal{Y}}(t)}) \;\; = \;\; \mathrm{vol}\{D_\alpha(F_{\boldsymbol{\mathcal{Y}}(t)})\} \, / \int_U \mathrm{vol}\{D_\alpha(F_{\boldsymbol{\mathcal{Y}}(u)})\}du.$$

The weight function $w_\alpha(t; F_{\boldsymbol{\mathcal{Y}}(t)})$ is proportional to the volume of the depth region at time point t. This implies that for regions where all curves nearly coincide the weight is small, heuristically, the order of the curves does not matter much here. For regions where the amplitude variability is large, there is a visual ordering of the curves, and the influence of those regions on the functional depth will be large.

While the definition of MFD needs a user-specified value of α, for many cases the definition does not depend on α. In general, the value of α is irrelevant when at each time point t the volumes of the depth regions are proportional to a fixed function of α. For many depth functions, this holds for elliptical symmetric distributions (Zuo and Serfling, 2000b) at each time point.

In Theorem 1 we show that the multivariate functional depth satisfies some key properties, adapted to a functional data context, that were put forward by Zuo and Serfling (2000a). All proofs as well as the conditions are relegated to the Appendix.

Theorem 1. *Assume that the depth function D satisfies the four properties listed in Zuo and Serfling (2000a), i.e. affine invariant, maximal at the center, monotone relative to the deepest point and vanishing at infinity. Then MFD, as defined in Definition 1, is a statistical depth function satisfying the following key properties:*
(i) Affine invariance (invariance w.r.t. the underlying coordinate system).

$$MFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) = MFD(\boldsymbol{A}\boldsymbol{X}_{(ct+d)} + \tilde{\boldsymbol{X}}_{(ct+d)}; F_{\boldsymbol{A}\boldsymbol{\mathcal{Y}}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}}, \alpha),$$

7

with $\boldsymbol{A}\boldsymbol{\mathcal{Y}}_{(ct+d)} + \tilde{\boldsymbol{X}}_{(ct+d)}$ the stochastic process $\{\boldsymbol{A}\boldsymbol{\mathcal{Y}}(\frac{s-d}{c}) + \tilde{\boldsymbol{X}}(\frac{s-d}{c}), s \in S = [cl + d, cu + d]\}$ for any constants $c \in \mathbb{R}_0$, $d \in \mathbb{R}$, any vector of functions $\tilde{\boldsymbol{X}} \in \mathcal{C}(U)^K$ and any matrix $\boldsymbol{A} \in \mathbb{R}^{K \times K}$ with $\det(\boldsymbol{A}) \neq 0$, and $\tilde{\boldsymbol{X}}_{(ct+d)}$ the curve $\{s, \tilde{\boldsymbol{X}}(\frac{s-d}{c})\}$, with $s \in S = [cl + d, cu + d]$.

(ii) Maximality at the center. $MFD(\boldsymbol{\Theta}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) = \sup_{\boldsymbol{X} \in \mathcal{C}(U)^K} MFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha)$, for any distribution $P_{\boldsymbol{\mathcal{Y}}}$ which has a uniquely defined point of symmetry $\boldsymbol{\Theta}$.

(iii) Monotonicity relative to the deepest point. $MFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) \leqslant MFD(\boldsymbol{\Theta} + a(\boldsymbol{X} - \boldsymbol{\Theta}); F_{\boldsymbol{\mathcal{Y}}}, \alpha)$, for any distribution $P_{\boldsymbol{\mathcal{Y}}}$ with a deepest point $\boldsymbol{\Theta}$ and for any $a \in [0, 1]$.

(iv) Vanishing at infinity. For $1 \leqslant k \leqslant K$ and for a series of curves $Y_{n,k}$ with $\lim_{n \to \infty} |X_{n,k}(t)| = \infty$ for almost all time points t in U: $\lim_{n \to \infty} MFD(X_{n,k}; F_{\mathcal{Y}}) = 0$.

In the original multivariate setting, the fourth property, 'vanishing at infinity', requires that for a vector $\boldsymbol{x} \in \mathbb{R}^K$ the depth of $\boldsymbol{x}$ should converge to 0 for $\|\boldsymbol{x}\| \to \infty$. When a curve behaves in accordance with the sample on the majority of the interval and converges to infinity near the border, one might not wish to attribute zero depth. The vanishing at infinity property for functional depth holds as stated in (iv).

2.2.2 Halfspace depth as a building block

In the remainder of the paper we mainly focus on halfspace depth (Tukey, 1975). The resulting MFD depth is called the *multivariate functional halfspace depth (MFHD)*. For $\mathcal{X}$ a random variable on $\mathbb{R}^K$ with cumulative distribution function $F_{\mathcal{X}}$ and a vector $\boldsymbol{x} \in \mathbb{R}^K$ the population halfspace depth (Tukey depth) is defined as

$$\mathrm{HD}(\boldsymbol{x}; F_{\mathcal{X}}) = \inf_{\boldsymbol{u} \in \mathbb{R}^K, \|\boldsymbol{u}\|=1} P(\boldsymbol{u}'\mathcal{X} \geqslant \boldsymbol{u}'\boldsymbol{x}). \tag{2}$$

It is well known that the halfspace depth regions are compact convex subsets of $\mathbb{R}^K$ (Rousseeuw and Ruts, 1999). Theorem 2 of Mizera and Volauf (2002) then guarantees through the uppersemicontinuity of the depth regions and the compactness of U,

8

together with the closedness and boundedness of the depth regions that the weight function is well defined. Moreover, at unimodal elliptic symmetric distributions, the contours of the depth regions coincide with density contours, which implies that the choice of α in MFHD becomes irrelevant (at least at the population level). We also choose HD because it satisfies the requirements of a building block for the functional depth as stated in Theorem 1. Consequently MFHD is affine invariant, maximal at the point (curve) of symmetry, monotone relative to the deepest point, and vanishing at infinity. An additional advantage of HD is its robustness with respect to outliers. The influence function of the HD of any multivariate point in $\mathbb{R}^K$ is bounded (Romanazzi, 2001) and the deepest point (Tukey median) has a positive breakdown value between $1/(K+1)$ and $1/3$ at absolutely continuous distributions (Chen and Tyler, 2002). Finally, fast algorithms exist for the computation of HD at multivariate data, as well as for the depth regions and the Tukey median (see Section 2.3 for details).

Theorem 2 states the existence of a deepest MFHD curve at absolutely continuous distributions $P_{\boldsymbol{\mathcal{Y}}}$ with a unique deepest point at each time point. As a result a MFHD median can be defined, see Definition 2.

Theorem 2. *Assume condition (A.1). Consider the curve $\boldsymbol{\Theta}$ which equals at each time point t the vector in $\mathbb{R}^K$ with maximum value of $HD(\cdot; F_{\boldsymbol{\mathcal{Y}}(t)})$. Then,*

1. *If the process $\boldsymbol{\mathcal{Y}}$ is such that $\mathcal{H} : U \times \mathbb{R}^K \to \mathbb{R} : (t, \boldsymbol{x}) \mapsto HD(\boldsymbol{x}, F_{\boldsymbol{\mathcal{Y}}(t)})$ is continuous, then $\boldsymbol{\Theta}$ is continuous: $\boldsymbol{\Theta} \in \mathcal{C}(U)^K$.*

2. *$\boldsymbol{\Theta}$ has maximal MFHD: for all $\boldsymbol{X} \in \mathcal{C}(U)^K$,*

$$MFHD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) \leqslant MFHD(\boldsymbol{\Theta}; F_{\boldsymbol{\mathcal{Y}}}, \alpha).$$

3. *Any curve $\tilde{\boldsymbol{\Theta}} \in \mathcal{C}(U)^K$ with maximal MFHD should have maximal HD at each time point t, i.e. $\tilde{\boldsymbol{\Theta}}(t) = \tilde{\boldsymbol{\theta}}_t$ with $HD(\tilde{\boldsymbol{\theta}}_t; F_{\boldsymbol{\mathcal{Y}}(t)}) = \max_{\boldsymbol{x} \in \mathbb{R}^K} HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)})$.*

Definition 2. *We define* $\mathbf{M}_{MFHD}(\boldsymbol{\mathcal{Y}}) = \boldsymbol{\Theta}$ *the MFHD median of* $\boldsymbol{\mathcal{Y}}$*, in which* $\boldsymbol{\Theta}(t)$ *is the vector in* $\mathbb{R}^K$ *with maximum value of* $HD(\cdot; F_{\boldsymbol{\mathcal{Y}}(t)})$.

Remark. When uniqueness of the deepest point at each time point is not assumed, we can define $\boldsymbol{\Theta}(t)$ as the center of mass of the set of points in $\mathbb{R}^K$ with maximum value of $HD(\cdot; F_{\boldsymbol{Y}(t)})$. In that case we obtain the upper hemicontinuity of the multifunction $t \mapsto \arg\max_{\boldsymbol{x}} HD(\boldsymbol{x}, F_{\boldsymbol{Y}(t)}) = \mathcal{G}(t)$. It might be possible to find conditions that guarantee that taking the center of mass of the set $\mathcal{G}(t) \subset \mathbb{R}^K$ is continuous as a function of $t \in U$, however, this would lead too far for the present purpose.

2.3 Finite sample definition

2.3.1 A general multivariate depth as building block

In practice one does not observe curves, but rather curve evaluations at a set of time points $t_1 < t_2 < \ldots < t_T$ in $U = [t_1, t_T]$, not necessarily equidistant.

Definition 3. *For a sample of multivariate curve observations* $\{\boldsymbol{Y}_1(t_j), \ldots, \boldsymbol{Y}_N(t_j); j = 1, \ldots, T\}$*, with at each time point* t *cdf* $F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t),N}$*, the sample multivariate functional depth at* $\boldsymbol{X} \in \mathcal{C}(U)^K$ *is defined by*

$$MFD_N(\boldsymbol{X}; \alpha) = \sum_{j=1}^{T} D(\boldsymbol{X}(t_j); F_{\boldsymbol{\mathcal{Y}}(t_j),N}) w_\alpha(t_j; F_{\boldsymbol{\mathcal{Y}}(t_j),N}) \tag{3}$$

with $w_\alpha(t_j; F_{\boldsymbol{\mathcal{Y}}(t_j),N})$

$$= vol\{D_\alpha(F_{\boldsymbol{\mathcal{Y}}(t_j),N})\}(t_{j+1} - t_{j-1}) / \{ \sum_{j=1}^{T} vol\{D_\alpha(F_{\boldsymbol{\mathcal{Y}}(t_j),N})\}(t_{j+1} - t_{j-1})\}.$$

The next theorem starts from the curves observed at a grid of time points and shows that the sample MFD is consistent for the population MFD under some conditions, when both N and T go to infinity.

Theorem 3. *(Consistency) Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$ be a sample with the same distribution as $\boldsymbol{\mathcal{Y}} \in C(U)^K$ with $E(\boldsymbol{\mathcal{Y}})$ finite, satisfying (A.1). We only observe these curves at time points $t_1 < t_2 < \ldots < t_T$ in U from a design generated as in (A.2). It holds that*

$$\sup_{\boldsymbol{X} \in C(U)^K} |MFD_N(\boldsymbol{X}; \alpha) - MFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha)| \to 0, \, a.s. \, P,$$

as $N \to \infty$ and $T \to \infty$ when the statistical depth D is such that it satisfies the conditions of Definition 1, as well as (i) $\sup_{\boldsymbol{x} \in \mathbb{R}^K} |D(\boldsymbol{x}; F_N) - D(\boldsymbol{x}; F)| \to 0, \, a.s. \, P,$ for $F_N \to F$ as $N \to \infty$ and (ii) $P(\{\boldsymbol{x} \in \mathbb{R}^K : D(\boldsymbol{x}; F) = \alpha\}) = 0.$

In the appendix we show that the finite sample MFD_N can be written as a population MFD applied to a set of interpolating continuous K-dimensional processes $\widetilde{\boldsymbol{\mathcal{Y}}}$, see (4), for which it holds that $\mathrm{MFD}_N(\boldsymbol{X}; \alpha) = \mathrm{MFD}(\boldsymbol{X}; F_{\widetilde{\boldsymbol{\mathcal{Y}}}, N}, \alpha)$.

2.3.2 Halfspace depth as a building block

As in the population case, we define the finite-sample multivariate functional halfspace depth (MFHD_N) as in Definition 3 with D the sample halfspace depth based on $\{\boldsymbol{Y}_1(t), \ldots, \boldsymbol{Y}_N(t)\}$ (Tukey, 1975),

$$\mathrm{HD}(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t), N}) = \frac{1}{N} \min_{\boldsymbol{u} \in \mathbb{R}^K, \|\boldsymbol{u}\|=1} \#\{\boldsymbol{Y}_n(t), n = 1, \ldots, N : \boldsymbol{u}'\boldsymbol{Y}_n(t) \geqslant \boldsymbol{u}'\boldsymbol{x}(t)\}.$$

The finite-sample Tukey median is defined as the center of gravity of the deepest depth region. The median curve of the sample $\{\boldsymbol{Y}_1(t_j), \ldots, \boldsymbol{Y}_N(t_j); j = 1, \ldots, T\}$ is defined as the Tukey median at each time point.

Exact computation of the MFHD_N can be done with fast algorithms for the halfspace depth up to dimension $K = 4$ (Bremner et al., 2008) and the depth contours up to dimension at least $K = 5$ (Hallin et al., 2010; Paindavaine and Šiman, 2012). In this paper we used the R-packages `depth` and `aplpack` which implement fast algorithms for bivariate and trivariate data (Rousseeuw and Ruts, 1996, 1998; Rousseeuw

and Struyf, 1998; Rousseeuw et al., 1999). Approximate halfspace depth in higher dimensions can be computed by means of the random Tukey depth (Cuesta-Albertos and Nieto-Reyes, 2008), but it is no longer affine invariant.

3 Example

We illustrate our new depth function on an industrial data set that produces one part during each cycle (De Ketelaere et al., 2011). The behavior of the cycle as monitored by an accelerometer provides a fingerprint of the cycle and, related, of the quality of the produced part. If a deviating acceleration signal occurs, the process owner should be warned. Figure 2(a) shows the acceleration signal of $N = 224$ parts measured during 120ms. Measurements are available every millisecond, hence the time signal ranges from $t_1 = 1$ up to $t_T = 120$. To augment this univariate data set, we could consider the derivatives of the curves as additional information, but in this example, we decided to use the integrated curves instead. As the velocity at time t_j, $V(t_j) = \int_{-\infty}^{t_j} A(t)dt$ with $A(t)$ the acceleration at time t, we approximated the velocity by $V(t_j) \approx V(t_{j-1}) + \{A(t_{j-1}) + A(t_j)\}/2$ starting with $V(t_1) = 0$. Note that the choice of the integration constant is not important here, due to the affine invariance of MFHD. The resulting velocity curves are depicted in Figure 2(b). Next, we performed a bivariate analysis on this data set, and computed the MFHD of all signals with $\alpha = 0.25$. The resulting depth values can be visualized by means of the so-called rainbow plot (Hyndman and Shang, 2010). It shows the curves colored according to their MFHD value. We first order the curves from maximal to minimal depth. Then we go from dark red for the deepest curve, to green for the curve with rank $N/2$, and move to dark blue for the curve with minimal depth. Such a representation of the data is in particular useful to visualize potential outliers, as they
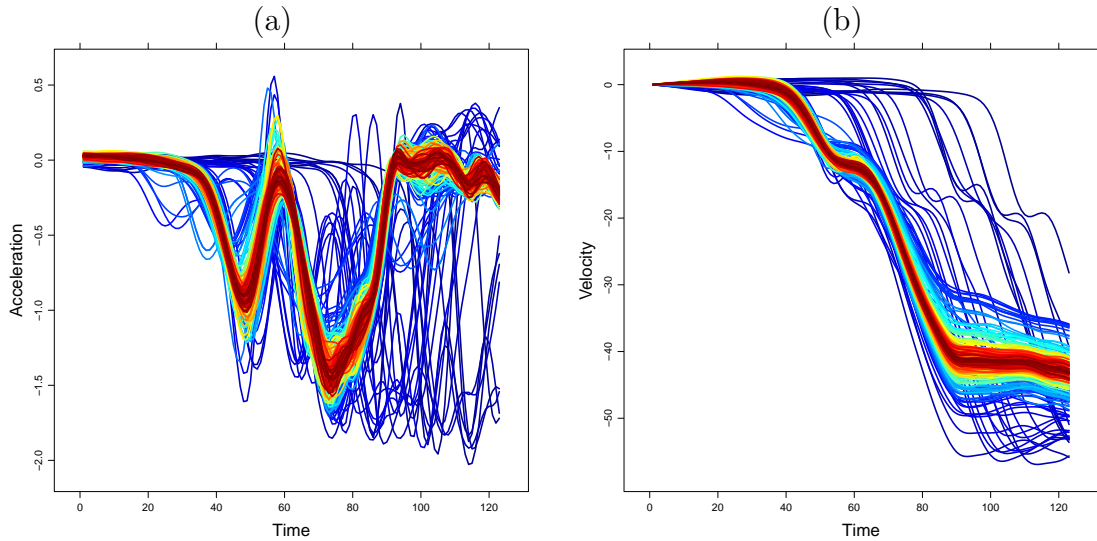
Figure 2: All signals colored according to their bivariate MFHD depth.

are expected to have a lower MFHD depth and consequently should be colored dark blue. In Figure 2 we see that the extreme outlying curves indeed are all colored dark blue, which is a confirmation that our depth measure assigns them a low depth value.

Computing the MFHD on the bivariate data $(A(t), V(t))$ also yields the median curves, printed in dark red on Figure 1 for the acceleration and velocity curves. We see that these median curves are not attracted by the outlying values at the end of the cycle. Also the acceleration estimates in the valleys around time points 50 and 75 are lower than those of the mean curve, illustrating the robustness of the median curve towards the upward contamination values in these regions. Note that the color of the mean and the median curve are chosen according to their MFHD. The mean curve has depth 0.217 which corresponds with the 0.75 quantile of all MFHD values. The median curve has depth 0.461, whereas the maximal depth among all observed curves is 0.317.

Apart from estimating the global pattern of the curves, we are interested in the variability of the curves. Our depth-based approach allows to visualize this dispersion

by means of the central regions, introduced in López-Pintado and Romo (2009). The β-central region consists of the band delimited by the $[n\beta]$ curves with highest depth. If we draw the 25%, 50% and 75% central regions, we obtain a representation of the data as in the enhanced functional boxplot of Sun and Genton (2011). Also here, we use the color scheme induced by the rainbow plot. Similarly, we can for example consider the 10%, 50% and 90% central regions. This yields Figures 3(a)–(b) for the acceleration curves and (c)–(d) for the velocity curves. Obviously, the 90% central regions contain outlying curves and hence increase the area of that central region.

Based on these central regions, we define for each univariate set of curves their *dispersion curves* $s_\beta(t)$ as the width of the β-central region at each t. These dispersion curves for $\beta = 0.1, 0.25, 0.5, 0.75$ and 0.9 are presented in Figure 4. They clearly expose the heteroscedasticity of both the acceleration and velocity signals, and they give an indication of the amount of outlying signals. Note that these curves are defined on each of the univariate curves, but the underlying computation of the central regions is based on the bivariate MFHD. The $s_{0.5}(t)$ dispersion curve can be considered as a kind of functional IQR, as already explained in Sun and Genton (2011). A related concept, the scale curve, is defined in López-Pintado et al. (2010). It measures the area of the central region for β ranging from 0 to 1, and could be considered here as well. Also the notions of β-trimmed mean and β-trimmed variance (as the mean and variance of all curves in the β-central region), see Fraiman and Muniz (2001), can be extended in a straightforward way.

4 Simulations

In this section we present four simulations settings each designed to illustrate a particular aspect of the behavior of MFHD. In all cases, we generate $N = 50$
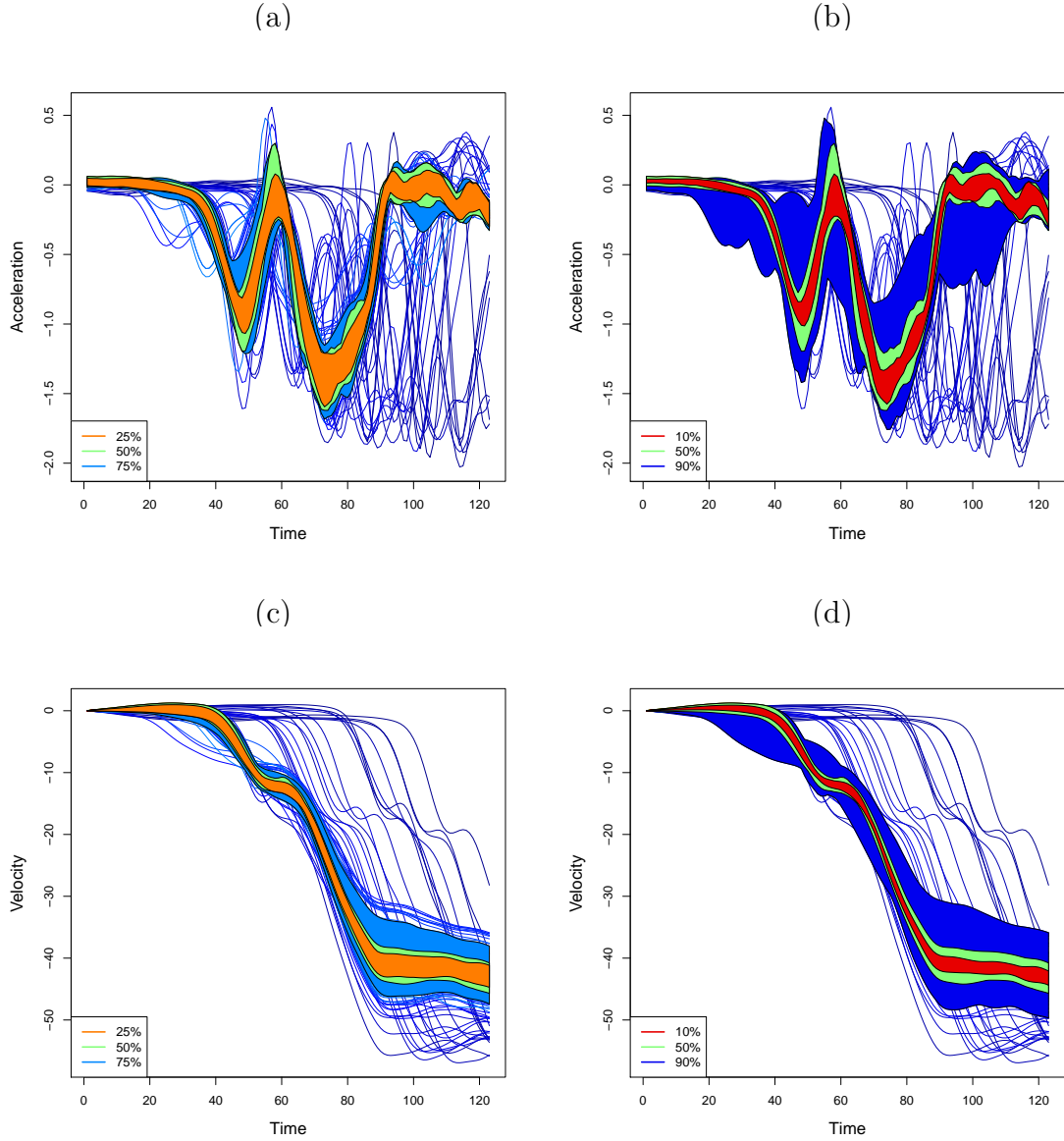
Figure 3: (a)–(b) Central regions for the acceleration curves and (c)–(d) for the velocity curves.

univariate curves $\{Y_1(t), Y_2(t), \ldots, Y_N(t)\}$ from a stochastic process $\mathcal{Y}$, denoted as the uncontaminated curves $\{Y(t)\}_N$. Then we replace five curves of $\{Y(t)\}_N$ with curves sampled from a contaminating stochastic process $\mathcal{Y}_\varepsilon$, yielding a data set $\{Y_1^\varepsilon(t), Y_2^\varepsilon(t), \ldots, Y_N^\varepsilon(t)\} = \{Y_\varepsilon(t)\}_N$ with 10% contamination. All curves are eval-
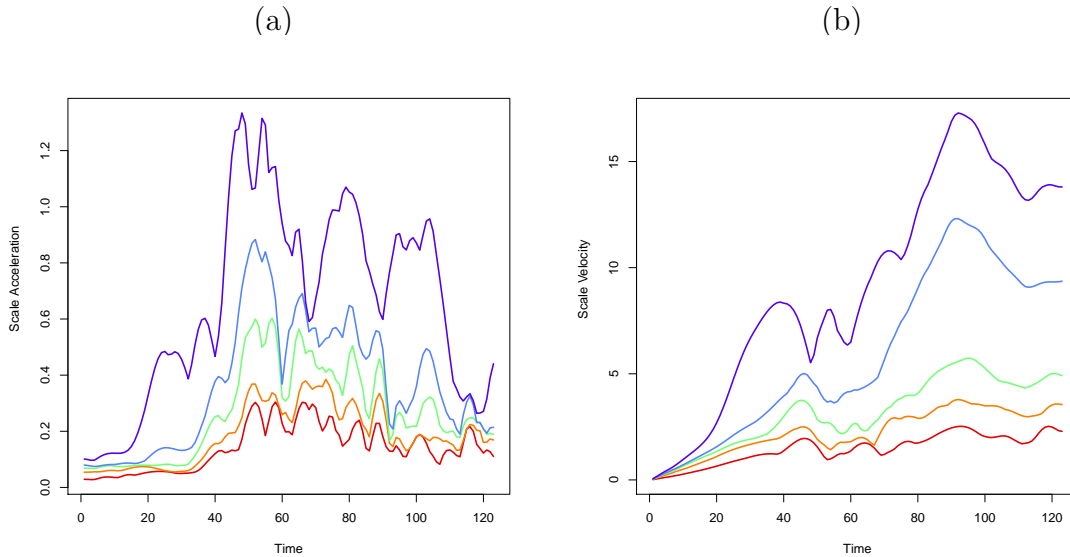
Figure 4: Dispersion curves for (a) the acceleration and (b) the velocity signals, for $\beta = 0.1$ (red), $\beta = 0.25$ (orange), $\beta = 0.5$ (green), $\beta = 0.75$ (light blue) and $\beta = 0.9$ (dark blue).

uated on a grid of $T = 100$ equispaced time points $t_1, t_2, \ldots, t_T$ in $[0, 2\pi]$. Each experiment was replicated 100 times. In a first set of simulations (Section 4.2) we consider the bivariate MFHD applied on the curves and their derivatives. We compare its behavior with several univariate functional depths and the bivariate random projection depth. In Section 4.3 we illustrate the advantage of using the warping functions (or a function thereof) as additional curves. First we describe how we evaluate the performance and robustness of the functional depths.

4.1 Evaluation criteria

ASE of the estimated central curve: the averaged squared scaled distance between the true and the estimated central curve,

$$\frac{1}{T} \sum_{j=1}^{T} \left(\frac{\widehat{m}_{Y_\varepsilon}(t_j) - m_{\mathcal{Y}}(t_j)}{s_{\mathcal{Y}}(t_j)} \right)^2,$$

16

where $m_{\mathcal{Y}}$ is the central curve of $\mathcal{Y}$, $\widehat{m}_{Y_\varepsilon}$ is the estimated central curve, and $s_{\mathcal{Y}}(t)$ is the interquartile range of $\mathcal{Y}(t)$.

ASE of the estimated dispersion curve: the average squared difference between the logarithm of the (0.5)-dispersion curves computed on the contaminated and the uncontaminated data,

$$\frac{1}{T} \sum_{j=1}^{T} \left(\log \left(\frac{s_{0.5}^{\varepsilon}(t_j)}{s_{0.5}(t_j)} \right) \right)^2,$$

where $s_{0.5}(t)$ is the width of the (0.5)-central region of $\{Y(t)\}_N$ as defined in Section 3. Analogously, $s_{0.5}^{\varepsilon}(t)$ is the dispersion curve computed from $\{Y_\varepsilon(t)\}_N$.

Normalized maximum depth of outliers. To have an easily comparable criterion, we normalize by dividing the maximum depth with the depth of the deepest curve. Formally, let $\mathrm{FD}_N(Y_n^\varepsilon, F_{\mathcal{Y}_\varepsilon, N})$ denote the (finite-sample) functional depth of any of the curves $Y_n^\varepsilon(t)$ from $\{Y_\varepsilon(t)\}_N$, and denote I_c the index set of the contaminated curves from $\{Y_\varepsilon(t)\}_N$. Then we consider $\max_{n \in I_c} \mathrm{FD}_N(Y_n^\varepsilon, F_{\mathcal{Y}_\varepsilon, N}) / \max_{n=1,\dots,N} \mathrm{FD}(Y_n^\varepsilon, F_{\mathcal{Y}_\varepsilon, N})$.

4.2 Simulations with curves and their derivatives

We generate three types of univariate curves, contaminated with 10% outlying curves, and we evaluate MFHD on the bivariate samples $\{(Y_\varepsilon(t), Y_\varepsilon'(t))\}_N$. We consider both $\alpha = 1/4$ and $\alpha = 1/6$, resulting in MFHD(1/4) and MFHD(1/6). Since the latter depth function works with larger depth regions, this could make it less robust. In both cases the estimated central curve $\widehat{m}_{Y_\varepsilon}(t)$ is the MFHD median, as in Definition 2.

We compare the behavior of MFHD on the bivariate samples, first, to three approaches applied on the univariate curves $\{Y_\varepsilon(t)\}$. (1) The cross-sectional average (CSA): $\widehat{m}_{Y_\varepsilon}(t) = \frac{1}{50} \sum_{n=1}^{50} Y_n^\varepsilon(t)$. The corresponding depth is the univariate Mahalanobis depth, with $\widehat{\sigma}_{Y_\varepsilon}(t)$ the cross-sectional standard deviation,

$$\mathrm{MD}(Y_n(t)) = \left\{ 1 + \left(\frac{Y_n(t) - \widehat{m}_{Y_\varepsilon}(t)}{\widehat{\sigma}_{Y_\varepsilon}(t)} \right)^2 \right\}^{-1}.$$

(2) The modified band depth (MBD) of López-Pintado and Romo (2009). This corresponds with $\mathrm{MFD}(Y_i, \mathcal{Y}_\varepsilon)$ as in (3) with the simplicial depth (Liu, 1990) as depth function D and with constant weight function $w_\alpha = 1/T$. The curve with largest MBD is considered as $\widehat{m}_{Y_\varepsilon}(t)$. We use the implementation provided in the R package `depthtools`. (3) The MFHD with $\alpha = 1/4$ applied on the univariate curves $\{Y_\varepsilon(t)\}_N$, which we denote by UFHD(1/4). Note that the deepest curve in this case corresponds with the cross-sectional median.

Next, we compare with the bivariate random projection depth (RPD) of Cuevas et al. (2007), using the default settings from the implementation in the R package `fda.usc` (Febrero-Bande and Oviedo de la Fuente, 2012). Here, the curves and their derivatives are projected on a random direction, yielding a bivariate sample on which the (bivariate) modal depth (Cuevas et al., 2006) can be computed for all observations. The RPD of a curve then corresponds with the average modal depth over 50 random projections. Note that this approach does not satisfy the affine equivariance property as stated in Theorem 1.

The functional derivatives are computed using B-splines using the default settings and the algorithms from the R-package `fda.usc`.

4.2.1 Simulation setting I: Shifted Outliers

This simulation setting illustrates the behavior of MFHD in cases where the true curves are homoscedastic and all derivative curves follow the same process. We generated curves of the form

$$Y_n^\varepsilon(t) = (1 - c_n)\{a_{1n}\sin(t) + a_{2n}\cos(t)\} + c_n\{a_{1n}\sin(t) + a_{2n}\cos(t) + \frac{1}{4}\},$$

where t is a grid of 100 equispaced values on $[0, 2\pi]$, c_n is 1 for 10% of the curves and 0 otherwise. The random coefficients a_{1n} and a_{2n} follow independent uniform
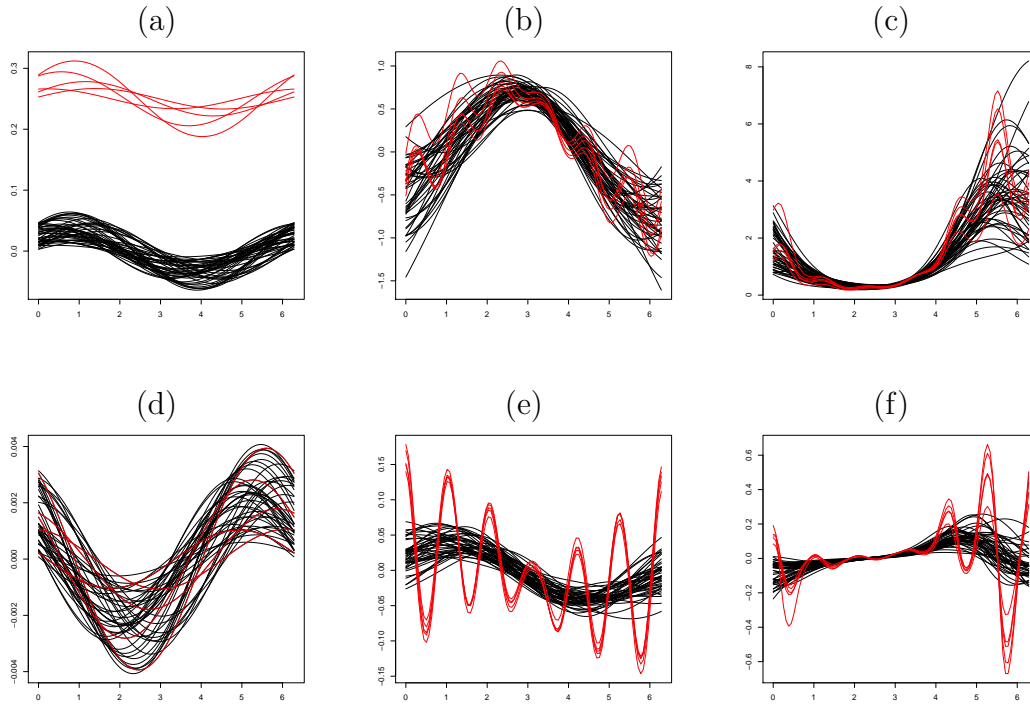
Figure 5: Simulation settings. Main curves (black) and outliers (red) for the curves (1st row) and their derivatives (2nd row) for (a),(d) the shifted outliers; (b),(e) the Gaussian processes; (c),(f) the log-normal processes.

distributions on $[0, 0.05]$. Figure 5(a) depicts the real (black) and outlying (red) curves curves and Figure 5(d) the corresponding derivatives.

The first panel in Figure 6 depicts, for each of the functional depth methods, the ASE of the central curves. CSA is highly influenced by the outlying curves, and RPD to a lesser extend. The middle panel in Figure 6 depicts the ASE of the dispersion curve. The effects of the outliers on the CSA estimate of dispersion $s_{0.5}(Y_n^\varepsilon(t))$ is more muted because the outlying curves are located too far to be included in the set of 25 curves with largest Mahalanobis depth. RPD contains some large values too. All other methods perform well on both criteria. The third panel in Figure 6 depicts the (normalized) maximum depth of outliers. Here again, CSA assigns a high depth to the outliers. Both univariate functional depths (MBD and UFHD) assign higher
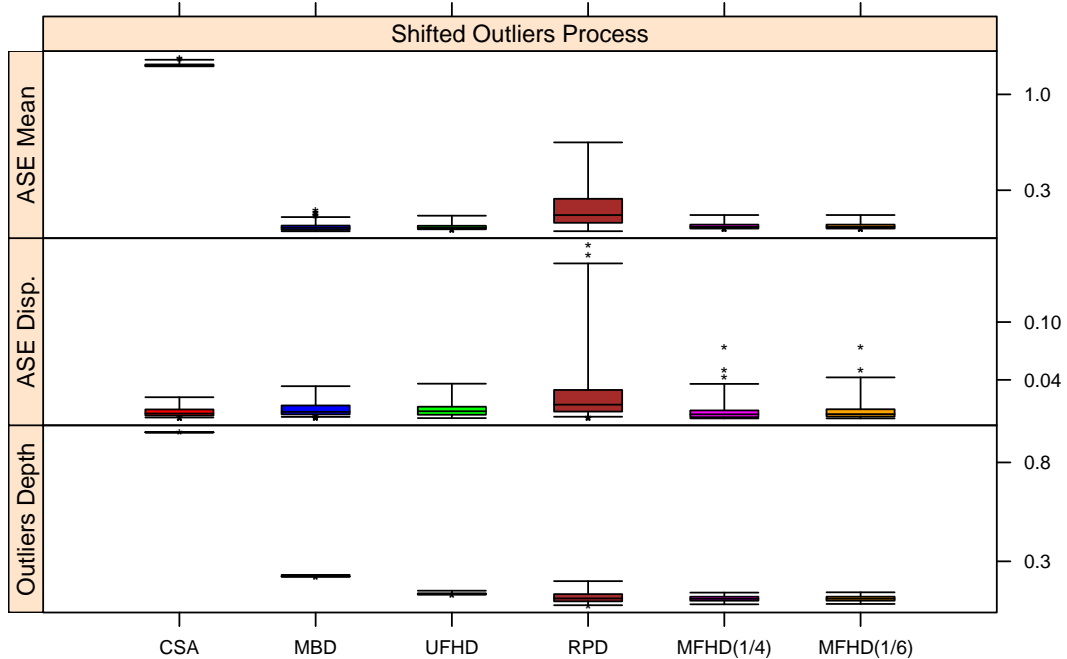
Figure 6: Setting I: shifted outliers. ASE of the estimated central curve (top), of the 0.5-dispersion curve (middle) and outlier depth (bottom).

depths to the outliers than the bivariate depth functions. The value of α for MFHD has a negligible effect on all three performance criteria.

4.2.2 Simulation setting II: Gaussian Process

The Gaussian processes $Y_n(t) \sim N(\mu(t), \Sigma(t))$ are generated as follows. Denote $\boldsymbol{x} = \{x_i\}_{i=1}^{20}$ 20 equidistant points on $[0, 2\pi]$. The covariance kernel of the x_i's is given by $K_{xx}(i, j) = \exp\left(-\frac{(x_j - x_i)^2}{2\delta^2}\right)$, with $\delta = 0.25$. For the time points $\boldsymbol{t} = \{t_j\}_{j=1}^{100}$ equidistant on $[0, 2\pi]$, we define K_{tx} and K_{tt} analogously. Then, the weight matrix for the mean of the process is $K^m = K_{tx}(K_{xx} + D)^{-1}$ where D is a diagonal matrix that directs the heteroscedasticity of the final process, $D = \text{Diag}_{i=1,\ldots,20}\{\min((\pi - x_i)^2, 1)\}$, so that the variance of the process is minimized at $t = \pi$. The mean function of the regular curves is taken as $\boldsymbol{\mu}(\boldsymbol{t}) = K^m(a_1 \sin(\boldsymbol{x}) + a_2 \cos(\boldsymbol{x}))$, where
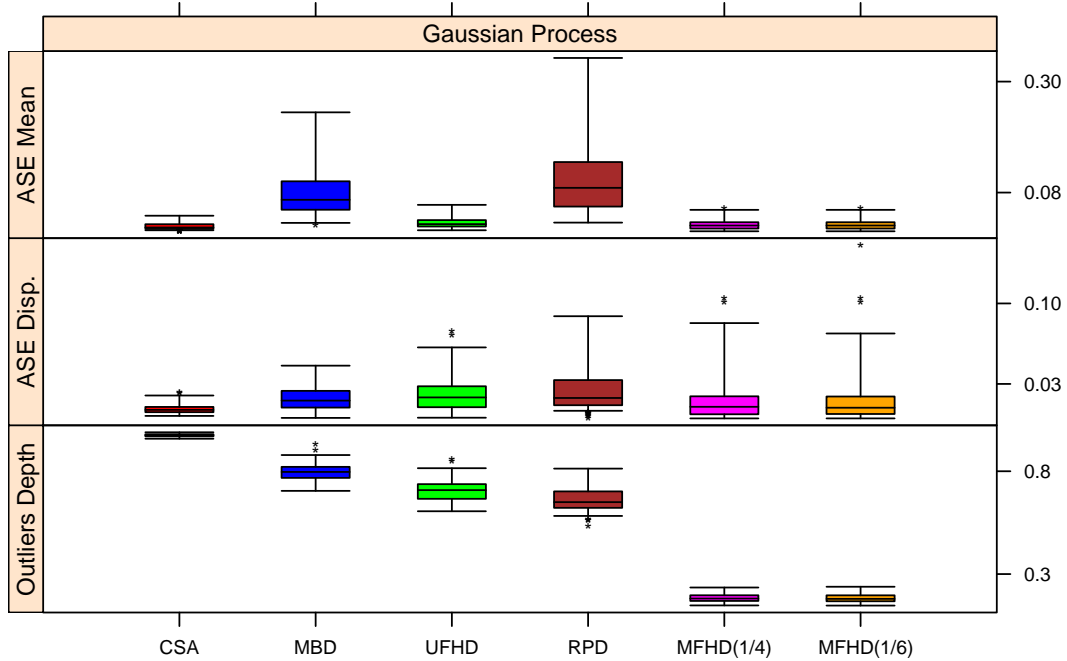
20

Figure 7: Setting II: Gaussian processes. ASE of the estimated central curve (top), of the 0.5-dispersion curve (middle) and outlier depth (bottom).

$a_1 \sim \mathcal{U}(-2, 2)$ and $a_2 \sim \mathcal{U}(-1, 1)$ are randomly generated. For the outlying curves we took $\boldsymbol{\mu}^*(\boldsymbol{t}) = K^m(\sin(6\boldsymbol{x}) + \boldsymbol{\mu}(\boldsymbol{x}))$. The covariance matrix of the process is given by $\Sigma = K_{tt} - K_{tx}(K_{xx} + D)^{-1}K'_{tx}$. The curves are mildly heteroscedastic, as can be seen from Figure 5(b).

This configuration was designed to penalize those estimators that do not use the information from the derivatives of the curves to assign depths. This is particularly visible in the third panel of Figure 7, where the MBD, CSA and UFHD are unable to detect the outlying curves. The outliers do not affect the CSA in terms of ASE of the central and the dispersion curves since the range of the response values is the same for all curves. Although RPD uses the derivatives, it does not perform well according to our three evaluation criteria.
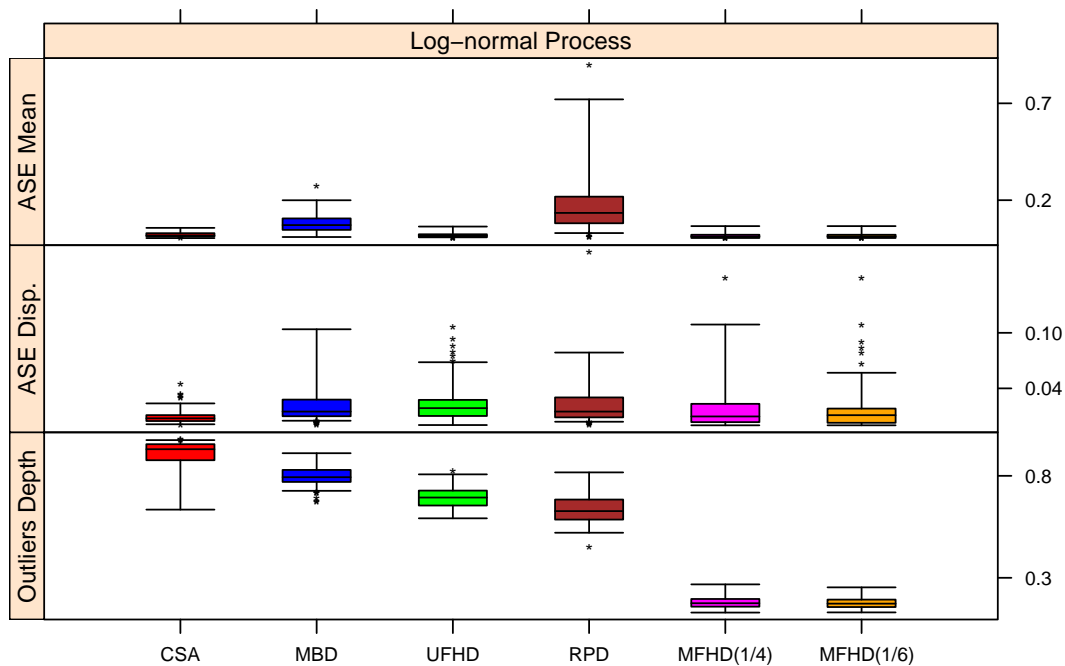
Figure 8: Setting III: log-normal processes. ASE of the estimated central curve (top), of the 0.5-dispersion curve (middle) and outlier depth (bottom).

4.2.3 Simulation setting III: Log-normal process

Highly heteroscedastic curves are obtained by substituting the normal distribution in setting II by a log-normal distribution. For a generated sample of curves and the corresponding derivatives, see Figure 5(c),(f). The difference with the Gaussian setting is most noticeable in terms of ASE of the central curve (top panel in Figure 8) where the scales are more than twice as large as compared to Figure 7). Again, RPD is not estimating the central curve well, and it does not assign low depths to the outlying curves. MFHD retains its good behavior.

4.3 Simulation with warped curves

Warping can make outlying curves more difficult to detect by pulling them towards the uncontaminated ones. See Figure 9 for an example where outlying curves are initially visible but are then hidden by the warping process. Here, using a bivariate approach can help with the ranking of the curves. For MFHD we compare two bivariate approaches. First, we create a bivariate sample of curves by using the warped curves together with the individual warping functions. Second, we use as a bivariate sample of curves the warped curves together with the derivatives of the warping functions. Adding the curves related to warping alleviates the information loss induced by the warping procedure.

For the warping functions, our simulation design follows that used in the first simulation setting of Arrabis-Gil and Romo (2012). The warping functions for the good curves are generated as explained on their page 405, formula (3.7). The inverse of

$$h_n(t) = \frac{\arctan(\beta_n(2t-1))}{2\arctan(\beta_n)} + 1/2, \ t \in [0,1]$$

with β_n equally spaced between 10 and 14 is used as warping function for the outliers. We use the same amplitude functions for all curves such that the difference in warping functions corresponds to original curves with different shapes and phases. To make the results comparable with those of simulation setting I, we use $Y_n(t) = a_{1i}\sin(t/(2\pi)) + a_{2i}\cos(t/(2\pi))$. A sample of curves and warped curves is depicted in Figure 9, together with the warping functions and their derivatives. For comparison we use two versions for UFHD: only the unwarped curves, and only the warped curves.

Figure 10 contains a summary of the performance criteria. As expected, once warped, the outlying curves have no influence on the estimation of the central (or dispersion curves) and this is visible in the first two panels of Figure 10 where UFHD has low MSE on both measures. At the same time, warping makes the curves with

23
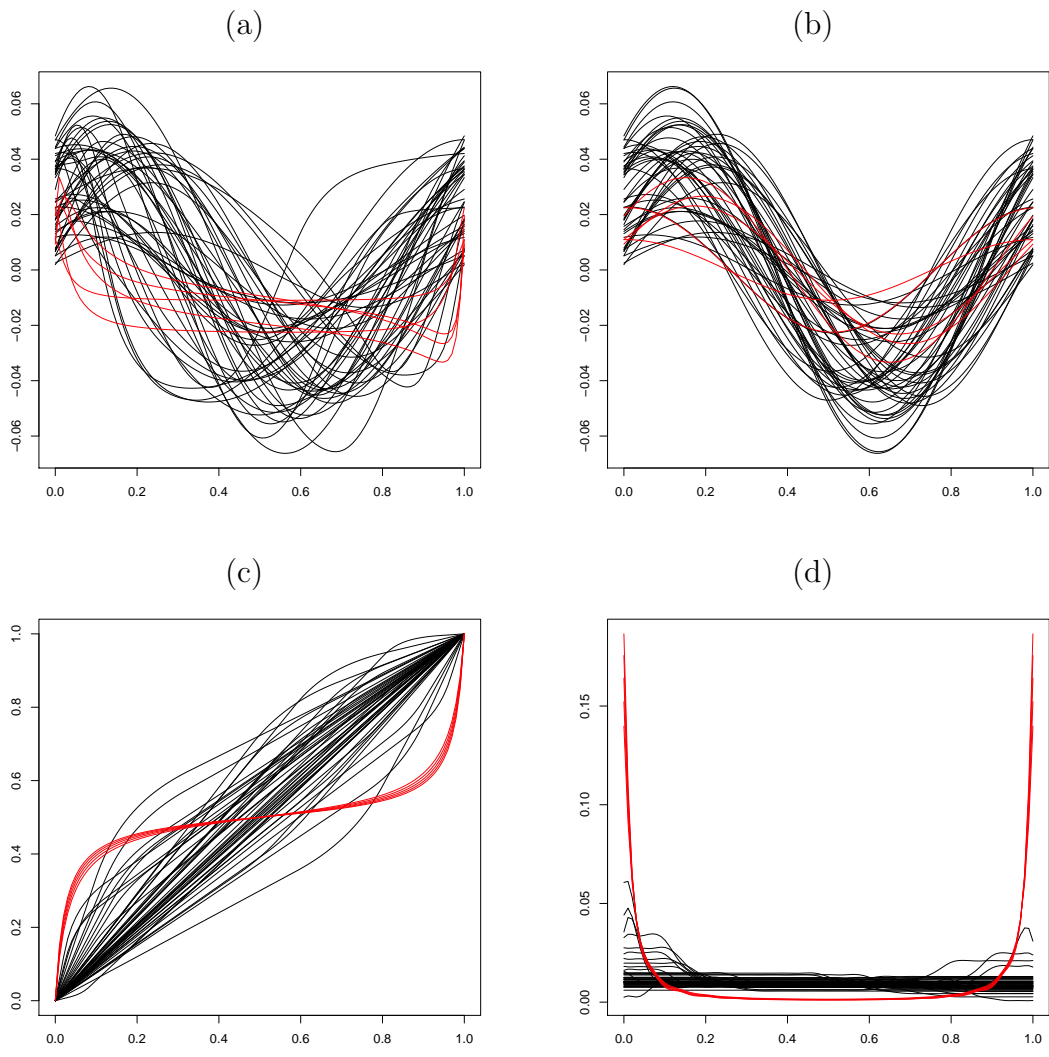
Figure 9: Simulation setting IV. (a) Original curves, (b) warped curves, (c) warping functions, (d) derivatives of the warping functions. Outlying curves are shown in red.

different shapes similar to the other curves, causing a poor behavior of UFHD on the warped curves in the third panel. Adding the warping functions, or their derivatives in a bivariate analysis completely addresses the information loss introduced by the warping procedure.
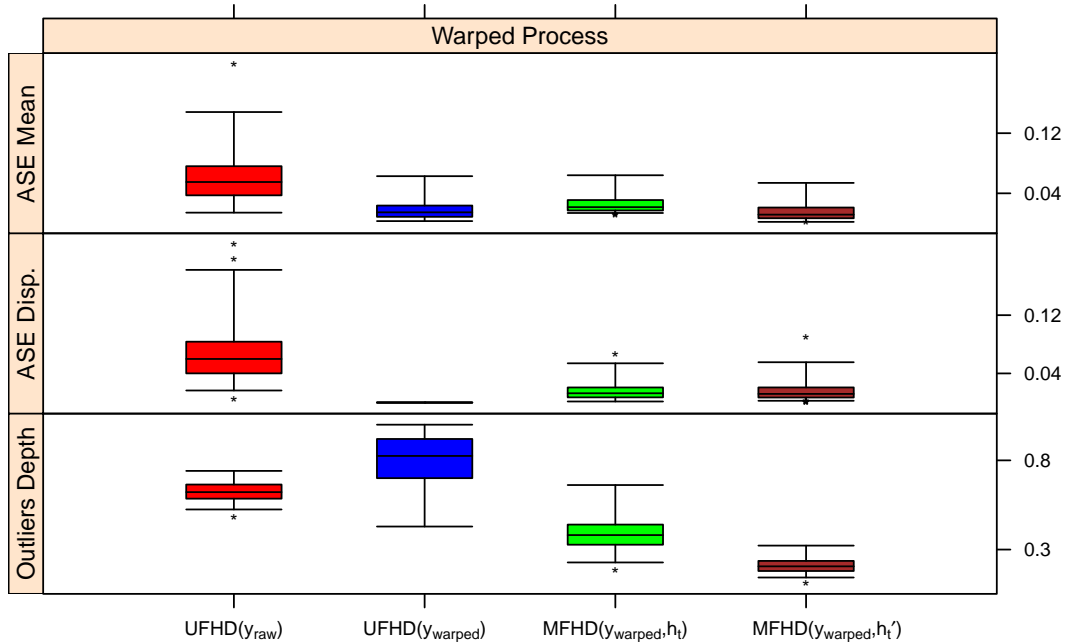
Figure 10: Setting IV. ASE of the estimated central curve (top) and the 0.5-dispersion curve (middle) and outlier depth (bottom). Using the original curves (y_{raw}), the warped curves (y_{warped}) and their warping functions (h_t) with derivatives h_t'.

5 Discussion

We have presented a new depth function for multivariate functional data (MFD), defined as a weighted average of the cross-sectional multivariate depths. It assigns a ranking to curves from the center outwards, whilst accounting for differences in amplitude. Shape and phase variation can be accommodated by including derivatives and/or warping functions. Interesting theoretical properties and computational advantages are achieved when using the multivariate halfspace depth, which leads to the MFHD depth function. The multivariate functional median curve can then be computed explicitly and estimates the central behavior of the curves. MFHD also allows to visualise and quantify the variability amongst the curves. Simulations have shown

the benefit of adding derivatives or warping information to univariate curves, and they have illustrated the better performance of MFHD compared with the bivariate random projection depth.

Depth functions are regularly used for the classification and clustering of multivariate data, see e.g. Ghosh and Chaudhuri (2005), Dutta and Ghosh (2011), Hubert and Van der Veeken (2010), Jörnsten (2004). This has been extended to the classification of functional data, as in Cuevas et al. (2007), López-Pintado and Romo (2006) and Hlubinka and Nagy (2012). It will be interesting to study the use of MFHD for the classification of multivariate curves, as well as for the classification of univariate curves augmented with their derivatives or warping functions. Some preliminary results indicate that this will indeed be beneficial (Slaets, 2011). Combining MFDH with the DD-plot of Li et al. (2012) is another interesting topic of further research. We will also investigate the possible use of MFHD for online quality control.

As illustrated on the data example and in the simulations, MFHD assigns lower depth to curves which deviate strongly from the majority of the curves. This robustness towards outliers is inherited from the halfspace depth which is applied at every time point. It is however well known that other depth functions, such as projection depth (Zuo, 2003), attain a higher breakdown value and thus could lead to a more robust MFD. Alternatively one could replace the integral in (1) (or equivalently the average in (3)) by an infimum as proposed in Mosler and Polyakova (2012). More theoretical and numerical studies are needed to compare these different depth functions.

A   Appendix

Some of this material might later be put in a supplementary materials file. We include it here to facilitate the review.

A.1 Assumptions

(A.1) The stochastic process $\boldsymbol{\mathcal{Y}}$ has continuous sample paths. That is, for each set B in the Borel-algebra of $\mathbb{R}^K$ the function $t \mapsto P_{\boldsymbol{\mathcal{Y}}(t)}(B) = P(\boldsymbol{\mathcal{Y}}(t) \in B)$ is continuous in t. The probability $P_{\boldsymbol{\mathcal{Y}}}$ is absolutely continuous and $P_{\boldsymbol{\mathcal{Y}}(t)}$ has a unique deepest point for every $t \in U$.

(A.2) Let the time points come from a design density f_T with $G(t) = \int_{-\infty}^{t} f_T(u)du$ such that $t_j = G^{-1}(\frac{j-1}{T-1})$ and that $U = [G^{-1}(0), G^{-1}(1)]$ is compact. We assume that f_T is differentiable and that $\inf_{t \in U} f_T(t) > 0$.

A.2 Proofs of Section 2

Lemma 1. *Consider a K-variate, real valued stochastic process of continuous functions $\boldsymbol{\mathcal{Y}}$ with continuous sample paths on a compact interval U. For an arbitrary $\boldsymbol{x} \in \mathbb{R}^K$, the function $t \mapsto HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)})$ is continuous.*

Proof. We show that $\forall \epsilon > 0, \exists \delta > 0$ such that for all t with $|t - t_1| < \delta : |HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)}) - HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t_1)})| < \epsilon$. Take $t_1 \in U$, $\boldsymbol{x} \in \mathbb{R}^K$ and $\epsilon > 0$ arbitrary but fixed. Since $\boldsymbol{\mathcal{Y}}(t)$ has continuous paths, there exists a $\delta > 0$ such that $\forall \boldsymbol{u} \in \mathbb{R}^K$ with $\|\boldsymbol{u}\| = 1$ and $\forall t \in U$ with $|t - t_1| < \delta : |P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t_1) \geq \boldsymbol{u}'\boldsymbol{x}) - P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t) \geq \boldsymbol{u}'\boldsymbol{x})| < \epsilon/2$, and thus also

$$\inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t_1) \geqslant \boldsymbol{u}'\boldsymbol{x}) \leqslant P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t) \geqslant \boldsymbol{u}'\boldsymbol{x}) + \epsilon/2$$

$$\inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t) \geqslant \boldsymbol{u}'\boldsymbol{x}) \leqslant P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t_1) \geqslant \boldsymbol{u}'\boldsymbol{x}) + \epsilon/2,$$

which implies that

$$\inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t_1) \geqslant \boldsymbol{u}'\boldsymbol{x}) \leqslant \inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t) \geqslant \boldsymbol{u}'\boldsymbol{x}) + \epsilon/2$$

$$\inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t) \geqslant \boldsymbol{u}'\boldsymbol{x}) \leqslant \inf_{\boldsymbol{u}} P(\boldsymbol{u}'\boldsymbol{\mathcal{Y}}(t_1) \geqslant \boldsymbol{u}'\boldsymbol{x}) + \epsilon/2.$$

And thus $|HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)}) - HD(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t_1)})| \leqslant \epsilon/2 < \epsilon$. $\qquad\square$

Lemma 2, below, is applicable for halfspace depth using Lemma 1 and Donoho and Gasko (1992, Lemma 6.1).

Lemma 2. *Consider a K-variate, real valued stochastic process of continuous functions $\boldsymbol{\mathcal{Y}}$ with continuous sample paths on a compact interval U. For a depth function D for which $t \mapsto D(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)})$ is continuous for all $\boldsymbol{x} \in \mathbb{R}^K$ and for which $\boldsymbol{x} \mapsto D(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)})$ is upper-semicontinuous for all $t \in U$, the following statement holds. Take $\boldsymbol{\Theta} \in \mathcal{C}(U)^K$. If there exists a $t_1 \in U$ and a $\boldsymbol{\theta}_{t_1} \in \mathbb{R}^K$ with $D(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{\mathcal{Y}}(t_1)}) < D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)})$ then there exists an interval $[t_1 - \delta, t_1 + \delta]$, $\delta > 0$ on which $D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) < D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t)})$.*

Proof. There exist values $\delta_1 > 0$ and $\delta_2 > 0$ such that $\forall a \in \mathbb{R}$ with $|a - D(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{\mathcal{Y}}(t_1)})| \leqslant \delta_1$ and $\forall b \in \mathbb{R}$ with $|b - D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)})| \leqslant \delta_2$: $a < b$. In particular: $D(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{\mathcal{Y}}(t_1)}) + \delta_1 < D(\boldsymbol{\theta}_{t_1}; P_{\boldsymbol{\mathcal{Y}}(t_1)}) - \delta_2$. Since $D(\cdot; F_{\boldsymbol{\mathcal{Y}}(t_1)})$ is upper semi-continuous on U, there exists a $\delta_3 > 0$: $D(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t_1)}) \leqslant D(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{\mathcal{Y}}(t_1)}) + \delta_1/2$ for $|\boldsymbol{x} - \boldsymbol{\Theta}(t_1)| < \delta_3$. Take $\boldsymbol{x} = \boldsymbol{\Theta}(t)$. Since $\boldsymbol{\Theta}$ is a continuous function it holds that for a $\delta_4 > 0$, $|\boldsymbol{\Theta}(t) - \boldsymbol{\Theta}(t_1)| < \delta_3$ when $|t - t_1| < \delta_4$. So far we thus have that $D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t_1)}) + \delta_1/2 < D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)}) - \delta_2$ for $|t - t_1| < \min(\delta_3, \delta_4)$. By the continuity of the function $t \mapsto D(\boldsymbol{x}; F_{\boldsymbol{\mathcal{Y}}(t)})$ there exists a $\delta_5 > 0$ such that $\forall t \in U$ with $|t - t_1| < \delta_5$: $D(\boldsymbol{\Theta}(t); P_{\boldsymbol{\mathcal{Y}}(t_1)}) + \delta_1/2 > D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)})$ and there exists a $\delta_6 > 0$, such that for $|t - t_1| < \delta_6$: $D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)}) \leqslant D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t)}) + \delta_2$. Together this implies that for t with $|t - t_1| < \min(\delta_3, \delta_4, \delta_5)$: $D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) < D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)}) - \delta_2$ and $\forall t$ with $|t - t_1| < \min(\delta_3, \delta_4, \delta_5, \delta_6) = \delta$: $D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) < D(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t)})$. $\square$

Proof of Theorem 1

Proof. (i) For a depth function D that is affine invariant and since $\text{vol}\{D_\alpha(\boldsymbol{\mathcal{Y}}(t))\} = |\det(\boldsymbol{A})| \cdot \text{vol}(D_\alpha\{\boldsymbol{A}\boldsymbol{\mathcal{Y}}(t) + \boldsymbol{X}(t))\}$, the affine invariance also holds for MFD.

(ii) For a center of symmetry $\boldsymbol{\Theta} \in \mathcal{C}(U)^K$ such that $P_{\boldsymbol{y}-\boldsymbol{\Theta}} = P_{\boldsymbol{\Theta}-\boldsymbol{y}}$ for a stochastic process $\boldsymbol{y}$, it holds that for each $t \in U$, $\boldsymbol{\Theta}(t)$ is a center of symmetry for $P_{\boldsymbol{y}(t)}$. Since D has maximality at the center, this ends the proof.

(iii) and (iv) are immediate from the assumptions on D. $\qquad\square$

Proof of Theorem 2

Proof. (1) We use Berge's maximum theorem (e.g. Abalo and Kostreva, 2005, Cor. 1). Define the function $\mathcal{H} : U \times \mathbb{R}^K \to \mathbb{R} : (t, x) \mapsto \mathrm{HD}(x, F_{\boldsymbol{y}(t)}) = \mathcal{H}(t, x)$. For $\mathcal{H}$ continuous and by uniqueness of the $\arg\max_{x \in \mathbb{R}^K} \mathrm{HD}(x, F_{\boldsymbol{y}(t)})$ for every $t \in U$, the continuity of Θ follows.

(2) This holds without assuming uniqueness of the deepest point. For any $t \in U$ there exists a point $\boldsymbol{x}_t$ in $\mathbb{R}^K$ with maximum depth for $P_{\boldsymbol{y}(t)}$ (Rousseeuw and Ruts, 1999, proposition 7). Denote $\mathrm{HD}(\mathbf{a}_t; F_{\boldsymbol{y}(t)}) = \alpha_t$. The set of all points with maximum depth at time t, $D_{\alpha_t}(F_{\boldsymbol{y}(t)})$, is convex and bounded (Rousseeuw and Ruts, 1999, propositions 1 and 5). The existence of a deepest curve is proven by defining $\boldsymbol{\Theta}(t) = \boldsymbol{\theta}_t$, the center of mass of $D_{\alpha_t}(F_{\boldsymbol{y}(t)})$. The convexity of $D_{\alpha_t}(F_{\boldsymbol{y}(t)})$ guarantees that $\boldsymbol{\theta}_t$ is a deepest point for all $t \in U$.

(3) We prove by contradiction that a deepest curve implies deepest points at every $t \in U$. We denote $\boldsymbol{\theta}_t = \boldsymbol{\Theta}_{\mathrm{HD}}(t)$ the deepest point of $P_{\boldsymbol{y}}$. Suppose that there exists a $t_1 \in U$ with $\mathrm{HD}(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{y}(t_1)}) < \mathrm{HD}(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{y}(t_1)})$. Because of Lemma 2: $\exists \delta > 0$ with for $|t - t_1| < \delta$: $\mathrm{HD}(\boldsymbol{\Theta}(t); F_{\boldsymbol{y}(t)}) < \mathrm{HD}(\boldsymbol{\theta}_{t_1}; F_{\boldsymbol{y}(t)}) < \mathrm{HD}(\boldsymbol{\theta}_t; F_{\boldsymbol{y}(t)})$. This implies that

$$
\begin{aligned}
\mathrm{MFHD}(\boldsymbol{\Theta}; F_{\boldsymbol{y}}, \alpha) &= \int_U \mathrm{HD}(\boldsymbol{\Theta}(t); F_{\boldsymbol{y}(t)}) w_\alpha(t; F_{\boldsymbol{y}}) dt \\
&< \int_{|t-t_1|>\delta} \mathrm{HD}(\boldsymbol{\Theta}_{\mathrm{HD}}(t); F_{\boldsymbol{y}(t)}) w_\alpha(t; F_{\boldsymbol{y}}) dt + \int_{t_1-\delta}^{t_1+\delta} \mathrm{HD}(\boldsymbol{\theta}_t; F_{\boldsymbol{y}(t)}) w_\alpha(t; F_{\boldsymbol{y}}) dt \\
&= \mathrm{MFHD}(\boldsymbol{\Theta}_{\mathrm{HD}}; F_{\boldsymbol{y}}, \alpha),
\end{aligned}
$$

which is in contradiction with the fact that $\boldsymbol{\Theta}$ is a deepest curve. $\qquad\square$

Finite sample MFD. Calculation of (3)

In order to apply Definition 1 to the curve observations at a grid of time points, we use linear interpolation on each interval $[t_j, (t_j + t_{j+1})/2]$ to connect the values $(t_j, Y_{nk}(t_j))$ with the average of the function values at time t_j in the middle of the time interval $((t_j + t_{j+1})/2, \bar{Y}_k(t_j) = N^{-1} \sum_{n=1}^{N} Y_{nk}(t_j))$, for $j = 1, \ldots, T-1$ and for each $k = 1, \ldots, K$. Similarly, for $j = 1, \ldots, T-1$, on the interval $[(t_j + t_{j+1})/2, t_{j+1}]$, linear interpolation is used to connect the values $((t_j + t_{j+1})/2, \bar{Y}_k(t_j))$ and $(t_{j+1}, Y_{nk}(t_{j+1}))$. This yields a sample of continuous K-dimensional processes $\widetilde{\boldsymbol{\mathcal{Y}}}_n$, $n = 1, \ldots, N$ on the interval $[t_1, t_T]$ of which the kth component ($k = 1, \ldots, K$) is defined by

$$\widetilde{\boldsymbol{\mathcal{Y}}}_{n,k}(t) = \begin{cases} Y_{nk}(t_j) \frac{t_j + t_{j+1} - 2t}{t_{j+1} - t_j} + \bar{Y}_k(t_j) \frac{2(t - t_j)}{t_{j+1} - t_j} & \text{for } t \in [t_j, (t_j + t_{j+1})/2] \\ -Y_{nk}(t_{j+1}) \frac{t_j + t_{j+1} - 2t}{t_{j+1} - t_j} - \bar{Y}_k(t_j) \frac{2(t - t_{j+1})}{t_{j+1} - t_j} & \text{for } t \in [(t_j + t_{j+1})/2, t_{j+1}]. \end{cases} \tag{4}$$

The empirical cumulative distribution function of this sample is denoted by $F_{\widetilde{\boldsymbol{\mathcal{Y}}},N}$. Note that the definition of the K-variate processes $\widetilde{\boldsymbol{\mathcal{Y}}}_n$ depends on both N and T. Using the definition of MFD on population level, see (1), with the processes $\widetilde{\boldsymbol{\mathcal{Y}}}_n$, and the affine invariance of the depth function gives Definition 3. Indeed,

$$\mathrm{MFD}(\boldsymbol{X}; F_{\widetilde{\boldsymbol{\mathcal{Y}}}_N}, \alpha) = \sum_{j=1}^{T-1} \left\{ \int_{t_j}^{(t_j + t_{j+1})/2} D(\boldsymbol{X}(t_j); F_{\boldsymbol{\mathcal{Y}}_N(t_j)}) \frac{|A_j(t)| \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(t_j)})}{\int_{t_1}^{t_T} \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(u)}) du} dt \right.$$
$$+ \left. \int_{(t_j + t_{j+1})/2}^{t_{j+1}} D(\boldsymbol{X}(t_{j+1}); F_{\boldsymbol{\mathcal{Y}}_N(t_{j+1})}) \frac{|(-1)^K A_j(t)| \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(t_{j+1})})}{\int_{t_1}^{t_T} \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(u)}) du} dt \right\},$$

where, with I_K the identity matrix with K rows,

$$A_j(t) = \mathrm{Det}(\frac{t_j + t_{j-1} - 2t}{t_{j+1} - t_j} I_K) = \left(\frac{t_j + t_{j-1} - 2t}{t_{j+1} - t_j} \right)^K,$$

implying that

$$\int_{t_j}^{(t_j + t_{j+1})/2} |A_j(t)| dt = \int_{(t_j + t_{j+1})/2}^{t_{j+1}} |(-1)^K A_j(t)| dt = \frac{t_{j+1} - t_j}{2(K+1)}.$$

The denominator in the weight function can similarly be written as

$$\int_{t_1}^{t_T} \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(u)}) du = \sum_{j=1}^{T} \mathrm{vol} D_\alpha(F_{\boldsymbol{\mathcal{Y}}_N(t_j)}) \frac{t_{j+1} - t_{j-1}}{2(K+1)}.$$

Proof of Theorem 3

Proof. On the space of curves in $C(U)^K$ we define the uniform distance $\rho(X, Y) = \sup_{t \in U} \|X(t) - Y(t)\|$, with $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^K$. We first show that the functions $\{\widetilde{\mathcal{Y}}_n : U \to \mathbb{R}^K : t \mapsto \widetilde{\mathcal{Y}}_{n,1}(t), \ldots, \widetilde{\mathcal{Y}}_{n,K}(t), n = 1, \ldots, N\}$ for $N \to \infty$ and $T \to \infty$ converge weakly to $\mathcal{Y}$. Since $E(\mathcal{Y})$ is finite, the law of large numbers implies that for each $t \in U$ and for $k = 1, \ldots, K$, $\bar{Y}_k(t) = N^{-1} \sum_{n=1}^{N} Y_{nk}(t) \to_{N \to \infty} E[\mathcal{Y}_k(t)]$, a.s.

By the design assumption (A.2), $|t_{j+1} - t_j| = |(G^{-1})'(\xi_j)|/(T-1) = c/(T-1)$, for a constant c and ξ_j in between t_j and t_{j+1}. For each $t \in U$ there is precisely one interval $[t_j, t_{j+1})$ that contains t and since the interpolating process agrees with the observed curve on the time points $t_1, \ldots, t_T$, it follows that for each value $n = 1, \ldots, N$,

$$0 \le \rho(\widetilde{\mathcal{Y}}_n, \boldsymbol{Y}_n) \le 2 \sup_{|s-t| \le c/(T-1)} \|\boldsymbol{Y}_n(t) - \boldsymbol{Y}_n(s)\| = 2 w_{\boldsymbol{Y}_n}(c/(T-1)),$$

with $w_{\boldsymbol{Y}}$ the modulus of continuity of $\boldsymbol{Y}$. Since U is compact and each function $\boldsymbol{Y}_n$ is continuous, this function is also uniformly continuous and thus $w_{\boldsymbol{Y}_n}(c/(T+1)) \to 0$ as $T \to \infty$. Since the sample of curves is i.i.d. and by Theorems 3.2 and 7.5 of Billingsley (1999), there is weak convergence of the interpolating processes $\widetilde{\boldsymbol{\mathcal{Y}}}_N$ to $\boldsymbol{\mathcal{Y}}$ as $N \to \infty$ and $T \to \infty$. It holds that

$$\sup_{\boldsymbol{X} \in C(U)^K} \left| \mathrm{MFD}(\boldsymbol{X}; F_{\widetilde{\boldsymbol{\mathcal{Y}}}, N}, \alpha) - \mathrm{MFD}(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \alpha) \right| \le \qquad (5)$$

$$\sup_{\boldsymbol{X} \in C(U)^K} \int_U \left| D(\boldsymbol{X}(t); F_{\widetilde{\boldsymbol{\mathcal{Y}}}, N}) - D(\boldsymbol{X}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) \right| dt + \int_U \left| w_\alpha(t; F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t), N}) - w_\alpha(t; F_{\boldsymbol{\mathcal{Y}}(t)}) \right| dt.$$

Under the stated assumptions, Theorem 4.1 of Zuo and Serfling (2000b) yields the a.s. convergence of the α-trimmed regions $D_\alpha(F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t), N})$ to $D_\alpha(F_{\boldsymbol{\mathcal{Y}}(t)})$, together with a nesting property such that for $0 < \epsilon < \alpha$, $D_\alpha(F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t), N}) \subset D_{\alpha - \epsilon}(F_{\boldsymbol{\mathcal{Y}}}(t))$ for all $t \in U$. By the dominated convergence, the right-hand side of (5) converges to 0 as $N, T \to \infty$. $\qquad \square$

References

Abalo, K. and Kostreva, M. (2005). Berge equilibrium: some recent results from fixed-point theorems. *Applied Mathematics and Computation*, 169(1):624–638.

Arrabis-Gil, A. and Romo, J. (2012). Robust depth-based estimation in the time warping model. *Biostatistics*, 13:398–414.

Berrendero, J., Justel, A., and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9):2619–2634.

Billingsley, P. (1999). *Convergence of Probability Measures*. John Wiley & Sons, Inc., second edition.

Bremner, D., Chen, D., Iacono, J., Langerman, S., and Morin, P. (2008). Output-sensitive algorithms for Tukey depth and related problems. *Statistics and Computing*, 18:259–266.

Chen, Z. and Tyler, T. (2002). The influence function and maximum bias of Tukey's median. *The Annals of Statistics*, 30:1737–1759.

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

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, 51(2):1063–1074.

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

De Ketelaere, B., Mertens, K., Mathijs, F., Diaza, D., and De Baerdemaeker, J. (2011). Nonstationarity in statistical process control - issues, cases, ideas. *Applied Stochastic Models in Business and Industry*, 27:367–376.

Donoho, D. and Gasko, M. (1992). Breakdown properties of location estimates

based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827.

Dutta, S. and Ghosh, A. (2011). On robust classification using projection depth. *Annals of the Institute of Statistical Mathematics*, 64:657–676.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package fda.usc. *Journal of Statistical Software*, 51(4):1–28.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer, New York.

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

Ghosh, A. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scandinavian Journal of Statistics. Theory and Applications*, 32(2):327–350.

Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *The Annals of Statistics*, 38(2):635–669.

Hlubinka, D. and Nagy, S. (2012). Functional data depth and classification. *Report.*

Hubert, M. and Van der Veeken, S. (2010). Robust classification for skewed data. *Advances in Data Analysis and Classification*, 4:239–254.

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

Jörnsten, R. (2004). Clustering and classification based on the L_1 data depth. *Journal of Multivariate Analysis*, 90(1):67–89.

Li, J., Cuesta-Albertos, J., and Liu, R. (2012). DD-classifier: nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107:737–753.

Liu, R. Y. (1990). On a notion of data depth based on random simplices. *The Annals*

of Statistics, 18(1):405–414.

López-Pintado, S. and Romo, J. (2006). Depth-based classification for functional data. In *Data depth: robust multivariate analysis, computational geometry and applications*, volume 72 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 103–119. Amer. Math. Soc., Providence, RI.

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

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

López-Pintado, S., Romo, J., and Torrente, A. (2010). Robust depth-based tools for the analysis of gene expression data. *Biostatistics*, 11:254–264.

Mizera, I. and Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *Journal of Multivariate Analysis*, 83(2):365–388.

Mosler, K. and Polyakova, Y. (2012). General notions of depth for functional data. *Report*.

Paindavaine, D. and Šiman, M. (2012). Computing multiple-output regression quantile regions. *Computational Statistics & Data Analysis*, 56:840–853.

Pigoli, D. and Sangalli, L. (2012). Wavelets in functional data analysis: estimation of multidimensional curves and their derivatives. *Computational Statistics & Data Analysis*, 56(6):1482–1498.

Ramsay, J. and Silverman, B. (2006). *Functional Data Analysis*. Springer, New York, 2 edition.

Romanazzi, M. (2001). Influence function of halfspace depth. *Journal of Multivariate Analysis*, 77:138–161.

Rousseeuw, P. and Ruts, I. (1996). Bivariate location depth. *Applied Statistics*, 45:516–526.

Rousseeuw, P. and Ruts, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica*, 8:827–839.

Rousseeuw, P. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49:213–244.

Rousseeuw, P., Ruts, I., and Tukey, J. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, 53:382–387.

Rousseeuw, P. and Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8:193–203.

Sangalli, L., Secchi, P., Vantini, S., and Veneziani, A. (2009). A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. *Journal of the American Statistical Association*, 104(485):37–48.

Slaets, L. (2011). *Analyzing phase and amplitude variation of functional data.* PhD thesis, KU Leuven.

Slaets, L., Claeskens, G., and Hubert, M. (2012). Phase and amplitude-based clustering for functional data. *Computational Statistics & Data Analysis*, 56:2360–2374.

Sun, Y. and Genton, M. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20(2):316–334.

Tukey, J. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531, Vancouver.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *The Annals of Statistics*, 31(5):1460–1490.

Zuo, Y. and Serfling, R. (2000a). General notions of statistical depth function. *The Annals of Statistics*, 28:461–482.

Zuo, Y. and Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *The Annals of Statistics*, 28(2):483–499.