

Flexible integrated functional depths

STANISLAV NAGY¹, SAMI HELANDER^{2,*}, GERMAIN VAN BEVER³,
LAURI VIITASAARI⁴ and PAULIINA ILMONEN^{2,†}

¹*Charles University, Prague, Czech Republic. E-mail: nagy@karlin.mff.cuni.cz*

²*Aalto University School of Science, Espoo, Finland.*

*E-mail: *sami.helander@aalto.fi; †pauliina.ilmonen@aalto.fi*

³*University of Namur, Namur, Belgium. E-mail: germain.vanbever@unamur.be*

⁴*Aalto University School of Business, Espoo, Finland. E-mail: lauri.viitasaari@aalto.fi*

This paper develops a new class of functional depths. A generic member of this class is coined J th order k th moment integrated depth. It is based on the distribution of the cross-sectional halfspace depth of a function in the marginal evaluations (in time) of the random process. Asymptotic properties of the proposed depths are provided: we show that they are uniformly consistent and satisfy an inequality related to the law of the iterated logarithm. Moreover, limiting distributions are derived under mild regularity assumptions. The versatility displayed by the new class of depths makes them particularly amenable for capturing important features of functional distributions. This is illustrated in supervised learning, where we show that the corresponding maximum depth classifiers outperform classical competitors.

Keywords: asymptotics; data depth; functional data analysis; integrated depths; supervised classification

1. Introduction

In this paper, we introduce a flexible class of depths for functional data. Such data, for which the observed units are random functions rather than random vectors, has become pervasive in many fields of modern research. Examples of functional data include growth curves in medical research, temperature, precipitation or water levels in meteorological studies, or electrode signals and fMRI data in brain studies. Many methods, originally developed for analyzing multivariate data, have been adapted to functional settings, see, for example, Ramsay and Silverman [43] and Ferraty and Vieu [15]. The richness of functional data typically makes it difficult to develop parametric models that generically and accurately capture the numerous features of the observations. This explains the large interest that nonparametric functional data analysis has received in the past decade.

In many statistical problems, a classical task is to assess whether an observation is a good representative of the population or not. In the multivariate setting, depth functions were introduced as a way to nonparametrically measure the centrality of a point with respect to an underlying distribution. Celebrated instances include the halfspace depth (Tukey [46]) and the simplicial depth (Liu [26]). An axiomatic approach to depth was developed in Zuo and Serfling [48]. Interestingly, depth notions allow to capture information not only about location but also about spread or shape of distributions. As a result, depth functions have been used in several inference problems including supervised classification (Ghosh and Chaudhuri [17], Li, Cuesta-Albertos and Liu [25]), testing for location (and scale) differences or diagnostics of non-normality (Liu, Parelius and Singh [27]). For these reasons, depth functions have been and are still under active research. Another example is the recent contribution by Chernozhukov et al. [6], which introduces a notion of Monge–Kantorovich depth. This depth generalizes the classical halfspace depth for one-dimensional or spherical distributions to general settings by using optimal transport.

The concept of data depth has been extended to functional settings. The first proposal in that direction was defined in Fraiman and Muniz [16], where the authors integrate univariate halfspace depth over the

functional values. Similar definitions, still of an integrated nature, include the modified band depth and half-region depth (López-Pintado and Romo [29,30]). An infimal approach, that considers the infimum of the marginal depth values, was adopted in Mosler [34], Mosler and Polyakova [35] and Narisetty and Nair [41].

Contrary to the multivariate case, assessing the centrality of an observation is not the sole objective in functional settings. Rather, recent proposals in this context aim at measuring typicality of an observation in a wider sense by, for example, taking the shape or roughness into account. Claeskens et al. [7] integrate multivariate marginal halfspace depth values of the joint vector consisting of the functions, their derivatives and/or warping functions. The objective of our paper is to introduce adaptive notions of functional depth, coined the J th order k th moment integrated depths, which automatically capture the relevant features of the functional distribution.

Studying theoretical properties of functional depths is known to be an ambitious task. Indeed, obtaining consistency already constitutes a challenge (Kuelbs and Zinn [23], Gijbels and Nagy [18]). Interestingly, the depth functions introduced in this paper are proven to be uniformly consistent. Moreover, a version of the law of the iterated logarithm is shown to hold. The asymptotic distributions of the sample versions are derived. This last result appears to be the first of its kind in the functional data depth literature.

The paper is organized as follows: Section 2 presents a motivating example that illustrates the benefits of using the adaptive functional depth. Section 3 provides a formal definition of the J th order k th moment integrated depth and details its theoretical properties. Section 4 is devoted to exploring its good finite sample properties and comparing its performance to other methods in classification. Finally, Section 5 discusses possible extensions and future topics of research, while the Appendix collects all the proofs from the paper.

2. Motivating example

Being able to assess the typicality of an observation is crucial in modern applications. For example, scientists worldwide routinely try to detect unusual patterns in climate, geological or astronomical data. Another example comes from finance, where rapidly detecting turbulence in the pattern of stock valuations is of importance. A more mundane task, although as difficult, involves sorting spam emails from regular correspondence. Note that, in this last case, the task does not amount to detecting outliers but flagging observations that do not share the same behavior as the rest.

The following example illustrates the need for ad hoc methods able to rank functional observations from the most typical to the most atypical. As discussed in the Introduction, depth methods provide such rankings.

The Adelaide electricity demand dataset¹ depicts the daily electricity demand in megawatts measured in Adelaide between July 1997 and March 2007. The dataset presented in Figure 1 contains 508 observations of Monday demand curves (in gray) to which 5 randomly chosen Sunday demand curves (in black) were added. All 513 curves are based on half-hourly measurements. The dataset will be further detailed in Section 4.

Detecting changes in the demand behavior is important for electricity producers and carriers as they need to be able to predict the consumption and adapt their production accordingly. Figure 1 illustrates the difficulty of the problem as, for example, the overall consumption on Sundays does not seem to differ significantly from that of Mondays. The consumption peaks, however, happen at different time points requiring the production to be adapted.

¹ Available at <https://rdrr.io/cran/fds/man/Electricitydemand.html>.

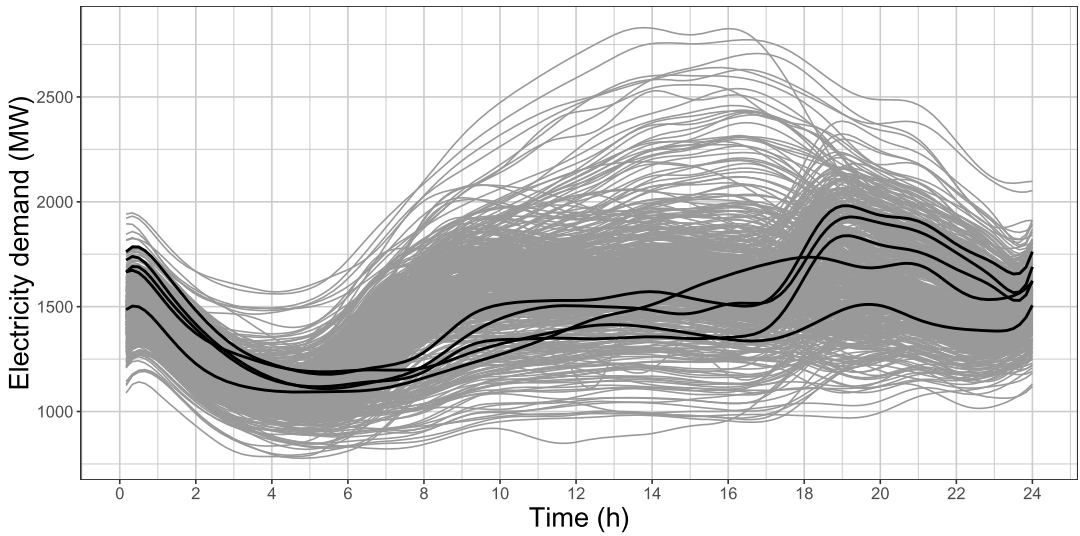


Figure 1. Monday electricity demands in Adelaide (gray) with 5 typical curves measured on Sundays (black).

Figure 2 displays boxplots of depth values of the 513 observations. In each of them, the depths of the Sunday curves are marked with asterisks. Each boxplot corresponds to a different functional depth. The functional depths considered here are the integrated halfspace depth (*ID*) from Fraiman and Muniz [16], the kernelized functional spatial depth (*KFSD*) from Sguera, Galeano and Lillo [44] (an extension of the celebrated functional spatial depth from Chakraborty and Chaudhuri [5]), the multivariate functional halfspace depth (*MFHD*) from Claeskens et al. [7] that computes the integrated bivariate depth of the functions and their derivatives, the extremal depth (*ED*) from Narisetty and Nair

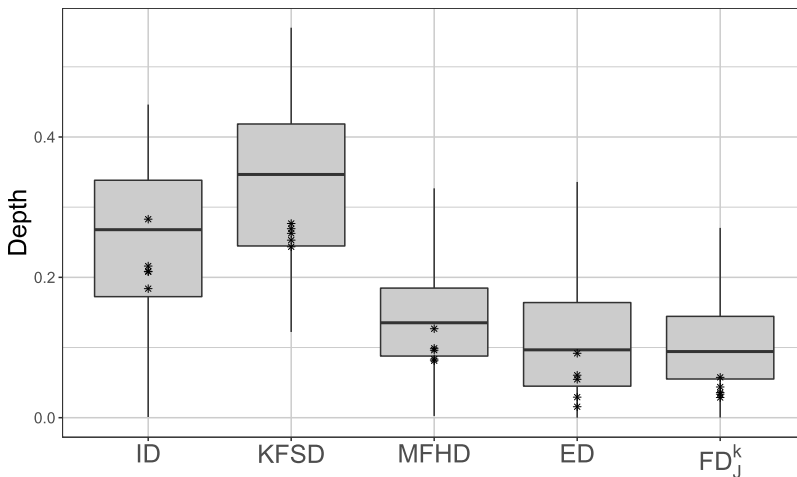


Figure 2. Boxplots of the depth values of the 513 curves described above for, from left to right, *ID*, *KFSD*, *MFHD*, *ED*, and FD_2^k . Sunday values are marked with asterisks.

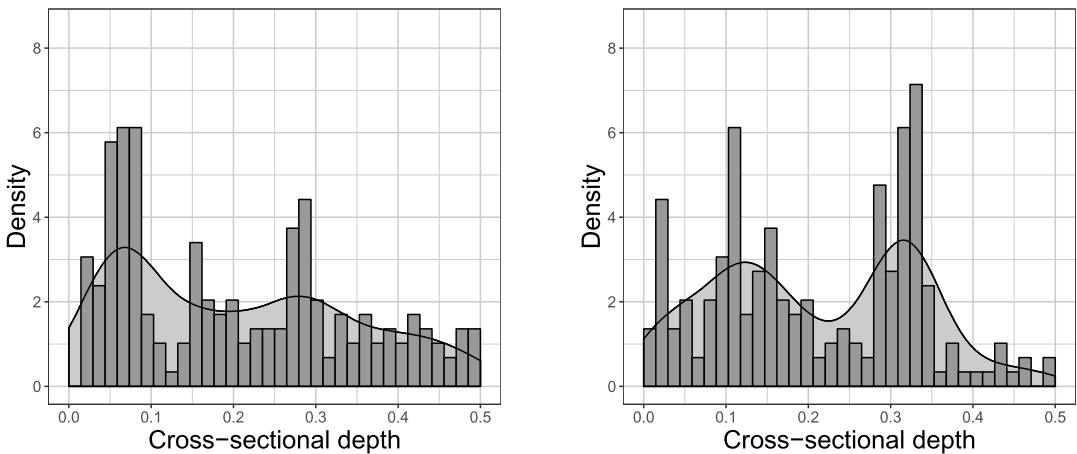


Figure 3. Histogram and kernel density estimate of marginal depth values for one randomly chosen Monday (left) and Sunday (right) curve.

[41], as well as the J th order k th moment integrated depth FD_J^k , proposed in this paper. The new depth is based on cross-sectional study of the observations, that is computing halfspace depths with respect to the marginal (in time) distributions. Depth FD_J^k is then based on characteristics of the distributions of these cross-sectional depths. We will later show that the parameter values used in this example ($J = 2$ and $k = -10$) allow to, in addition to location, take the shape of the functions into account.

The boxplots illustrate the good performance of FD_J^k in flagging atypical days. Indeed, the Sunday curves are assigned depth values that are comparatively lower than those of the main bulk of the data. The only other depth able to capture, however only two of, the added curves is ED . Note that the general location of the boxplots is not relevant in this context since the interest lies in the ranking that the depths provide rather than their values themselves. The ability of FD_J^k to detect unusual curves is based on the fact that it considers the entire distribution of vectors of marginal (in time) depth values.

Studying vectors of marginals and other moments is more informative than comparing univariate average measures. Figure 3 illustrates this by showing histograms and kernel density estimates of the cross-sectional univariate depth values (taken at 200 equispaced time points) of one randomly chosen Monday curve and one randomly chosen Sunday curve. The average depth values in Figure 3 are, for the Monday and the Sunday curve, respectively, 0.202 and 0.211, so that it is very challenging to distinguish them based only on the integrated depth (see the ID boxplot in Figure 2). Similarly, infimal depth assigns value close to 0 to both curves. Both densities exhibit two peaks and the shapes of the distributions are rather similar. This suggests to also consider joint distributions of the depth values at different time points to include information about the shape of the curves.

This data example highlights the need to construct a depth that considers not only the univariate depth distribution of cross-sectional values but also their joint marginal distribution as well as other moments. The following section formally defines such depth, that is, the J th order k th moment integrated depth FD_J^k . Moreover, it provides its asymptotic properties.

3. Moment-based functional depths

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability space on which every random quantity is defined. Throughout, let $\mathcal{P}(A)$ denote the set of all Borel probability distributions on the measurable set A . In the sequel,

$D(\cdot; P)$ denotes the celebrated halfspace depth function under distribution P (Tukey [46]). Depending on the context, it either refers to the univariate version

$$D(x; P) = \min \left\{ F_P(x), 1 - \lim_{y \rightarrow x^-} F_P(y) \right\} \quad \text{for } x \in \mathbb{R} \text{ and } P \in \mathcal{P}(\mathbb{R}),$$

where $F_P(x) = \mathbb{P}(X \leq x)$ denotes the value of the (univariate) cumulative distribution function of $X \sim P$ at $x \in \mathbb{R}$, or it refers to the multivariate version

$$D(\cdot; P) : \mathbb{R}^J \rightarrow [0, 1] : x \mapsto D(x; P) = \inf_{u \in S^{J-1}} \mathbb{P}(u^\top X \geq u^\top x), \tag{3.1}$$

where X has distribution $P \in \mathcal{P}(\mathbb{R}^J)$ and $S^{J-1} = \{u \in \mathbb{R}^J : u^\top u = 1\}$ denotes the hypersphere in \mathbb{R}^J . For continuous distributions P , the depth (3.1) attains values only in $[0, 1/2]$.

Turning now to the functional setting, let $x \in L^2([0, 1])$ and $X \sim P \in \mathcal{P}(L^2([0, 1]))$, where the latter means that $X : [0, 1] \times \Omega \rightarrow \mathbb{R}$ is jointly measurable and, for each $\omega \in \Omega$, $X(\cdot, \omega) \equiv X(\cdot) \in L^2([0, 1])$. Throughout, $P_t \in \mathcal{P}(\mathbb{R})$ denotes the distribution of $X(t)$, the value of X at point t . For $\mathbf{t} = (t_1, \dots, t_J)^\top$, the distribution of $X(\mathbf{t}) \equiv (X(t_1), \dots, X(t_J))^\top$ is denoted by $P_{\mathbf{t}} \in \mathcal{P}(\mathbb{R}^J)$. In the sequel x denotes a generic element of either \mathbb{R} , \mathbb{R}^J or $L^2([0, 1])$. The context will make clear to which we are referring. Note that it is customary in functional data analysis to assume integrability of x . In this context, however, instead of integrability, it would be sufficient to assume that x is Borel measurable (Nagy [36]).

We next define the J th order k th moment integrated depth. The parameter k relates to the moments of the distribution of cross-sectional depth values that are under study. This allows to adaptively select features of the distribution that best distinguish the observations. The parameter J defines the number of time points that are considered simultaneously. This allows to use not only location to assess depth but also to take the shape properties of the functional observations into account. This is similar in spirit to the approach taken in Nagy, Gijbels and Hlubinka [38] and allows to consider changes in the slope or convexity of the observations without requiring existence of derivatives.

Definition 1. Let $J = 1, 2, \dots$ and $|k| \geq 1$. For $P \in \mathcal{P}(L^2([0, 1]))$, the J th order k th moment integrated depth with respect to P is the function

$$FD_J^k(\cdot; P) : L^2([0, 1]) \rightarrow [0, 1]$$

$$: x \mapsto FD_J^k(x; P) = \left(\int_{[0,1]^J} (D(x(\mathbf{t}); P_{\mathbf{t}}) + 1/2)^k d\mathbf{t} \right)^{1/k} - 1/2.$$

For $x \in L^2([0, 1])$ and $k = \pm\infty$, define

$$FD_J^\infty(x; P) = \lim_{k \rightarrow \infty} FD_J^k(x; P) \quad \text{and} \quad FD_J^{-\infty}(x; P) = \lim_{k \rightarrow -\infty} FD_J^k(x; P).$$

Note that the choice to use (a function of) the halfspace depth in the integrated depth definition provided above is of a theoretical nature. Naturally, any choice of multivariate depth can be considered and would provide a corresponding functional integrated version (for example, considering simplicial depth and $k = J = 1$ would yield the particular case of a modified band depth from López-Pintado and Romo [29]). We will, however, restrict to the halfspace case, which provides interesting theoretical properties, that we now detail. See also the discussion at the end of Section 3.1.

The limits in the previous definition always exist due to the boundedness of the halfspace depth and the monotonicity of the L^p norm. Note that the expression inside the integral involves $D(x(\mathbf{t}); P_{\mathbf{t}}) +$

$1/2$. The constant was added in order to guarantee the existence of the integral for $k \leq -1$. The constant is then subtracted to ensure that $FD_J^k(x; P) \in [0, 1]$.

For $k = 1$ and $J = 1$, Definition 1 coincides with the usual integrated depth for functional data (Fraiman and Muniz [16]), that is, the average cross-sectional depth of a functional value of x with respect to the corresponding marginal distribution of X .

Writing $\|\cdot\|_k$ for the $L^k([0, 1]^J)$ -norm of a function defined on $[0, 1]^J$, Definition 1 rewrites, for $1 \leq k < \infty$,

$$FD_J^k(x; P) = \|D(x(\cdot); P) + 1/2\|_k - 1/2.$$

Similarly, for $-\infty < k \leq -1$ and $l = -k = |k|$,

$$FD_J^k(x; P) = 1/\left\|\frac{1}{D(x(\cdot); P) + 1/2}\right\|_l - \frac{1}{2}. \quad (3.2)$$

For $k = \infty$, $FD_J^k(\cdot; P)$ is the essential supremum of the cross-sectional depth function. Similarly,

$$\begin{aligned} FD_J^{-\infty}(x; P) &= \lim_{k \rightarrow -\infty} FD_J^k(x; P) = \lim_{l \rightarrow \infty} 1/\left\|\frac{1}{D(x(\cdot); P) + 1/2}\right\|_l - \frac{1}{2} \\ &= \left(\operatorname{ess\,sup}_{t \in [0, 1]^J} \frac{1}{D(x(t); P_t) + 1/2}\right)^{-1} - \frac{1}{2} \\ &= \left(\frac{1}{\operatorname{ess\,inf}_{t \in [0, 1]^J} D(x(t); P_t) + 1/2}\right)^{-1} - \frac{1}{2} = \operatorname{ess\,inf}_{t \in [0, 1]^J} D(x(t); P_t). \end{aligned}$$

Therefore, for $k \rightarrow -\infty$, the moment integrated depths approximate the (essential) infimum of the marginal depths of functional values. This approach is closely related to the technique called the infimal (or extremal) depth for functional data, considered by Mosler [34], Mosler and Polyakova [35] and Narisetty and Nair [41].

Altogether, with the moment extensions of the integrated depth for random functions, we obtain a versatile collection of depth functionals that encompasses both the original integrated and infimal depths, and many other intermediate functionals. The idea of considering the whole integrand function in the computation of the integrated depth (by taking its moments) can be seen to be related to the approach pursued by Narisetty and Nair [41]. Contrary to their suggestion of using the minimal value, or quantiles, of the depth distribution, we show in the sequel that there are several advantages in using moments. Indeed, depths defined using moments are (i) always uniquely defined, (ii) universally (and uniformly) consistent (which is not the case for infimal depths, see Gijbels and Nagy [18]), and (iii) well defined also for discontinuous functional data. The choice of k and J in practice will be driven by the application considered.

3.1. Sample depth consistency

The sample version $FD_J^k(x; P_n)$ is obtained by plugging in $P_n \in \mathcal{P}(L^2([0, 1]))$ – the empirical measure of a random sample from P of size n . Theorem 1 below shows that, remarkably, the sample depths introduced above are, for all $1 \leq |k| < \infty$ and all $J = 1, 2, \dots$, universally consistent over the whole space $L^2([0, 1])$. Results of this type are not available for infimal depths, or other non-integrated versions of depths for infinite-dimensional data (López-Pintado and Romo [29], López-Pintado and Romo [30], Chakraborty and Chaudhuri [5], Sguera, Galeano and Lillo [44]). It is important to mention that

in the limit cases $k = \pm\infty$ the consistency may fail to hold true, as can be seen by modification of the example using the infimal depth in $L^2([0, 1])$ of Nagy and Ferraty [37], p. 99.

Theorem 1. For all $P \in \mathcal{P}(L^2([0, 1]))$, $J = 1, 2, \dots$, and $1 \leq |k| < \infty$,

$$\sup_{x \in L^2([0,1])} |FD_J^k(x; P_n) - FD_J^k(x; P)| \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0$$

and

$$\sup_{x \in L^2([0,1])} |FD_J^k(x; P_n) - FD_J^k(x; P)| = \mathcal{O}_P(n^{-1/2}).$$

The next theorem states an upper bound for the functional depth $FD_J^k(\cdot; P)$ in the spirit of the law of the iterated logarithm. It is the first result of this type for any functional depth.

Theorem 2. For all $P \in \mathcal{P}(L^2([0, 1]))$, $J = 1, 2, \dots$, and $1 \leq |k| < \infty$, the following holds true almost surely

$$\limsup_{n \rightarrow \infty} \sup_{x \in L^2([0,1])} |FD_J^k(x; P_n) - FD_J^k(x; P)| \sqrt{\frac{2n}{\log \log n}} \leq C_k,$$

where $C_k = 1$ for $1 \leq k < \infty$ and $C_k = 9$ for $-\infty < k \leq -1$.

Theorem 2 gives a uniform version of an asymptotic result that is intermediate between the uniform consistency of Theorem 1, and the central limit theorem of Theorem 3 given in the next section.

Note that, in Theorems 1 and 2, it is not crucial that the finite-dimensional depth D used in the definition of FD_J^k is the halfspace depth. In fact, Theorem 1 still applies (after obvious modifications) for integrated depths based on any uniformly consistent finite-dimensional depth that satisfies appropriate uniform rates of convergence. Similarly, for any depth D that follows a bound based on the law of the iterated logarithm, an analogue to Theorem 2 holds true.

3.2. Distributional asymptotics

Next, we study the asymptotic distribution of $FD_J^k(\cdot; P)$ and adapt a uniform weak convergence result for the halfspace depth process given by Massé [33], Theorem 2.1, to the functional setting. To do so, we introduce the following notations. For $u \in S^{J-1}$ (recall that S^{J-1} denotes the hypersphere in \mathbb{R}^J) and $x \in \mathbb{R}^J$, write

$$H[x, u] = \{y \in \mathbb{R}^J : u^\top y \geq u^\top x\}$$

for the closed halfspace whose boundary passes through x with inner normal u and

$$\mathcal{H} = \{H[x, u] : x \in \mathbb{R}^J, u \in S^{J-1}\}$$

the set of all halfspaces in \mathbb{R}^J . For $Q \in \mathcal{P}(\mathbb{R}^J)$ and $x \in \mathbb{R}^J$, a halfspace $H[x, v] \in \mathcal{H}$ is called a minimal halfspace at $x \in \mathbb{R}^J$ if $Q(H[x, v]) = D(x; Q)$. Let

$$V(x) = \{v \in S^{J-1} : Q(H[x, v]) = D(x; Q)\}$$

denote the set of all inner normals of minimal halfspaces of $Q \in \mathcal{P}(\mathbb{R}^J)$ at x .

Further, we need some terminology and results from the theory of empirical processes. All the facts used hereafter about these processes can be found in van der Vaart and Wellner [47], or Dudley [12]. For $H \in \mathcal{H}$, $Q \in \mathcal{P}(\mathbb{R}^J)$ and Q_n the empirical measure of a random sample of size n from Q , consider the empirical process $v_n(H) = \sqrt{n}(Q_n(H) - Q(H))$. Each realization of this process is an element of the space $\ell^\infty(\mathcal{H})$ of bounded maps from \mathcal{H} to \mathbb{R} equipped with the uniform norm. The process v_n is known to converge weakly (in the sense of Hoffmann-Jørgensen) to a Q -Brownian bridge v_Q – a centered Gaussian process indexed by \mathcal{H} that is a tight, Borel-measurable map into $\ell^\infty(\mathcal{H})$ and whose covariance function is given by $E(v_Q(H)v_Q(H')) = Q(H \cap H') - Q(H)Q(H')$.

As shown by Massé [33], Theorem 2.1, under some technical assumptions, the halfspace depth process $\sqrt{n}(D(x; Q_n) - D(x; Q))$, indexed by a subset of some $x \in \mathbb{R}^J$, converges weakly to a process given by

$$\mathcal{J}(v_Q)(x) = \inf_{v \in V(x)} v_Q(H[x, v]). \tag{3.3}$$

In particular, if for $x \in \mathbb{R}^J$ the set $V(x)$ is a singleton (i.e., there exists a single minimal halfspace at x), the limit (3.3) is Gaussian. Otherwise the limit still exists, but may not be Gaussian. We extend this result to $FD_J^k(\cdot; P)$.

Theorem 3. *Let $P \in \mathcal{P}(L^2([0, 1]))$ be such that*

$$P_t(\partial H) = 0 \quad \text{for all } H \in \mathcal{H} \text{ and all } t \in [0, 1]^J. \tag{S}$$

Then, for any $x \in L^2([0, 1])$,

$$\begin{aligned} &\sqrt{n}(FD_J^k(x; P_n) - FD_J^k(x; P)) \\ &\xrightarrow[n \rightarrow \infty]{\mathcal{D}} (FD_J^k(x; P) + 1/2)^{1-k} \int_{[0,1]^J} (D(x(t); P_t) + 1/2)^{k-1} \mathcal{J}(v_{P_t})(x(t)) dt, \end{aligned}$$

where the process $\mathcal{J}(v_{P_t})(x(t))$ in the integrand on the right hand side is (3.3) for measure P_t , evaluated at $x(t)$.

The smoothness condition (S) is satisfied if each marginal distribution P_t has a density.

For the simplest case $k = 1$ and $x \in L^2([0, 1])$ with unique minimal halfspaces at each $t \in [0, 1]^J$, Theorem 3 reduces to an asymptotic normality result. Its proof follows from that of Theorem 3 in a straightforward way.

Corollary 4. *Under the assumptions of Theorem 3, suppose further that $X \sim P$, and let $H_t = H[x(t), v(t)] \in \mathcal{H}$ be the unique minimal halfspace of P_t at $x(t)$, for all $t \in [0, 1]^J$. Denote for $t_1, t_2 \in [0, 1]^J$*

$$\begin{aligned} \alpha(t_1) &= P(X(t_1) \in H_{t_1}) = D(x(t_1); P_{t_1}), \\ \beta(t_1, t_2) &= P([X(t_1) \in H_{t_1}] \cap [X(t_2) \in H_{t_2}]). \end{aligned}$$

Then the sequence $\sqrt{n}(FD_J^k(x; P_n) - FD_J^k(x; P))$ is distributed asymptotically as a centered Gaussian random variable with variance

$$c \int_{[0,1]^J} \int_{[0,1]^J} ((\alpha(t_1) + 1/2)(\alpha(t_2) + 1/2))^{k-1} (\beta(t_1, t_2) - \alpha(t_1)\alpha(t_2)) dt_1 dt_2,$$

where $c = (FD_J^k(x; P) + 1/2)^{2(1-k)}$.

Corollary 4 holds, for instance, for P a non-degenerate Gaussian process on $[0, 1]$, and x different from the mean of P except on a finite subset of $[0, 1]$.

Note that the distributional asymptotics shown in the results above are the first of their kind in the literature. The only comparable result is Cuevas and Fraiman [9], Theorem 4. There, the asymptotic distribution such as that in Corollary 4 is obtained for an integrated depth similar to our moment integrated depth with $k = 1$ and $J = 1$. However, Cuevas and Fraiman [9] use a univariate version of the simplicial depth in the definition of the integrated depth, instead of the halfspace depth. The simplicial depth is known not to satisfy some desirable properties that are expected from a depth, such as quasi-concavity (Zuo and Serfling [48]). Additionally, the derivation of Cuevas and Fraiman [9], Theorem 4, is based on the representation of the univariate simplicial depth with respect to $Q \in \mathcal{P}(\mathbb{R})$ in terms of the distribution function of Q . That result cannot be extended to $J > 1$.

4. Depth based classification of real data

Supervised learning is a statistical problem that arises in several application areas. Depth methods have proved useful in such contexts. In this section we consider supervised classification of functional data and provide several data examples which allow to assess the flexibility of the depth functions defined in Section 3. In particular, we illustrate the ability of FD_j^k to capture features of the functional observations that are relevant to conduct classification.

There is an extensive literature related to supervised functional classification. References include Hall, Poskitt and Presnell [20] or Ferraty and Vieu [14]. More recently, Biau, Bunea and Wegkamp [1] and Cérou and Guyader [4] studied k -nearest neighbors methods and their properties in infinite dimensional spaces. Also, Delaigle and Hall [10] considered classifiers based on dimension reduction through either partial least squares or functional principal components projections.

Depth based methods have also been considered for functional classification. Ghosh and Chaudhuri [17] introduced the celebrated maximal depth approach for depth-based classification which allocates a new observation to the population in which it has the highest depth. This allows to, generically, turn any depth function into a classifier. Extensions of the max-depth classifier include Cuevas, Febrero and Fraiman [8] (to the functional context) and Li, Cuesta-Albertos and Liu [25] (for adaptive depth-depth plot classification). It has now become a common practice to compare depth functions based on their empirical classification properties.

We next consider a supervised functional classification exercise, in which we compare the following classifiers:

1. the k -nearest neighbors procedure (kNN) from Biau, Bunea and Wegkamp [1] using leave-one-out cross-validation on k .
2. the Delaigle and Hall [10] classifiers based on principal components (PC) or partial least squares (PLS).
3. the max-depth classifiers based on (i) Claeskens et al. [7] multivariate functional halfspace depth ($MFHD$) applied to the functions and their derivatives, (ii) Sguera, Galeano and Lillo [44] kernelized functional spatial depth ($KFSD$) with a Gaussian kernel and automatic bandwidth selection provided by the authors, and (iii) the J th order k th moment integrated depth FD_j^k . For this last depth, different values of $J \in \{1, 2, 3\}$ were considered together with a leave-one-out cross-validation on the parameter $k \in \{k' \in \mathbb{Z}: 1 \leq |k'| \leq 20\} \cup \{\pm\infty\}$.

Note that max-depth classifiers based on the modified band depth (López-Pintado and Romo [29]) or the integrated depth from Fraiman and Muniz [16] are not considered here as these depths are only special cases of FD_j^k . The max-depth classifier in 3.(iii) therefore readily improves on classification conducted with these depths.

Computation of FD_J^k was performed as follows. Functions in each dataset were measured on a corresponding time index set. For $J = 1$ the cross-sectional depths were computed over the entire index set. For $J = 2$ and $J = 3$ the depths were computed over 5000 uniformly sampled J -variate vectors of time indices. As the halfspace depth is invariant to permutations of the chosen J time points, only permutationally distinct vectors were considered in the computation. The integrals appearing in the definition of FD_J^k were approximated using averages over the selected (J -variate) index set. For $k = +\infty$ (resp. $k = -\infty$), FD_J^k was approximated as the maximum depth (resp. minimum depth) over the index set.

For each dataset, leave-one-out misclassification rates of the presented classifiers are reported. Note that the multivariate maximum depth classifier traditionally breaks ties at random, according to weights depending on group sizes. We will adopt the same convention for the functional data considered here. While ties are considerably less frequent in this case (due to integration), they still may appear. This will be the case, for example, if the function has depth 0 with respect to all groups.

Additionally, an X -fold out-of-sample classification scheme was conducted for FD_J^k in each dataset. In the out-of-sample classification scheme, each dataset was split at random into X parts of equal sizes. One at a time, each subsample was treated as the training set, and the remaining $X - 1$ parts as the testing set. In the training set, a range of different values of k were tested. The k that minimized the leave-one-out misclassification rate on the training set was then used to evaluate the leave-one-out misclassification rates on the testing set. The choice of X was guided by the total sample size of each dataset, resulting in $X = 10$ for the Adelaide electricity consumption dataset and $X = 5$ for the Tecator and the Australian weather datasets. Notably, there were no significant differences in neither the chosen values of k nor the resulting misclassification rates of FD_J^k between the leave-one-out and the out-of-sample classification schemes. Therefore, in the following, we focus on the leave-one-out misclassification rates as the performance criterion of the methods in supervised classification. Additional results of the simulation study can be found in the on-line supplemental document (Nagy et al. [40]).

4.1. Description of the datasets considered

4.1.0.1. *Australian weather dataset.* The Australian rainfall dataset² (see Delaigle and Hall [10] and Li et al. [24] for details on data preprocessing) depicts the average daily precipitation at 190 weather stations in Australia between 1840 and 1990.

Two classification problems are considered for this dataset. The North–South grouping divides the data into two distinct rainfall patterns. The observations in the Northern population exhibit a “tropical” pattern with most of the precipitation during summer months, whereas the Southern population has most of its rainfall during cooler winter months. The West–East grouping is associated with changes in the amplitude of the precipitation: larger changes in the West and more stable precipitation in the East due to the presence of the Great Dividing Range. Both partitions are described in Li et al. [24]. The latter problem is often considered more difficult as the relevant features for classification are related to the shape of the observations rather than their location. The geographical groupings of the weather stations and the data are presented in Figure 4.

4.1.0.2. *Tecator spectrometric dataset.* The Tecator spectrometric dataset³ (described in Borggaard and Thodberg [3]) depicts the light wavelength absorbance of 215 finely chopped pieces of meat analyzed by the Tecator Infratec spectrometer. The absorbance of each meat sample was measured at

²Available at <https://rda.ucar.edu/datasets/ds482.1/>.

³Available at <http://lib.stat.cmu.edu/datasets/tecator>.

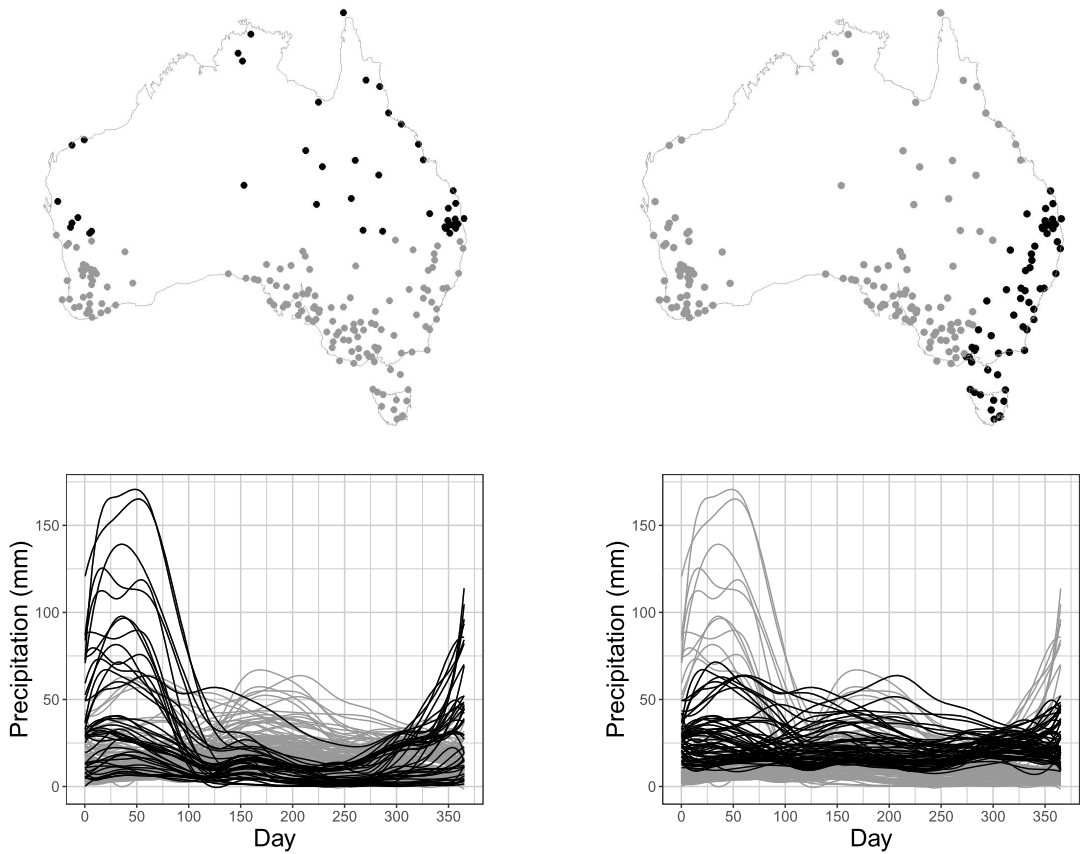


Figure 4. Geographical locations of the Australian weather stations with the North–South (upper left) and West–East (upper right) groupings highlighted in gray and black. Bottom row: corresponding yearly rainfall data in matching colors.

100 uniformly spaced discretization points over the near-infrared wavelength range of 850–1050 nm. The classification problem for this dataset has been considered regularly in the functional data analysis literature (see, for example, Ferraty and Vieu [15]). Meat samples have been separated into two groups based on the fat content of the sample.

The spectrometric data is presented in Figure 5. The classical grouping – 138 observations with low fat content ($\leq 20\%$) vs 77 observations with high fat content ($> 20\%$) – is represented by using gray and black lines, respectively.

Classification of the spectrometric curves is difficult as the two groups overlap, exhibiting very similar smoothness properties and spread of the vertical location shift. However, the observations differ in curvature.

4.1.0.3. Adelaide electricity consumption dataset. The Adelaide electricity demand dataset⁴ (see Magnano, Boland and Hyndman [31] and references therein) contains the daily electricity demand in megawatts measured in Adelaide between 1997 and 2007. For each week day, there are 508 curves,

⁴Available at <https://rdrr.io/cran/fds/man/Electricitydemand.html>.

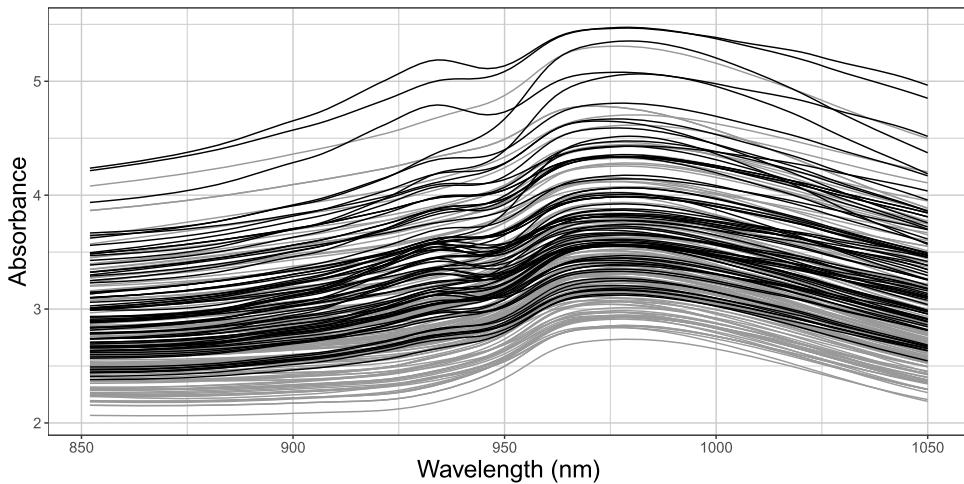


Figure 5. Tecator light wavelength absorbance data of 215 samples of meat. Observations with less (resp. more) than 20% of fat content are in gray (resp. black).

based on half-hourly measurements. As for the Australian dataset, the Adelaide data was smoothed using B -splines. For the purpose of this analysis, two business days, Monday and Tuesday, and the two weekend days, Saturday and Sunday, were chosen, and four leave-one-out classification schemes were conducted pairing both business days with both weekend days. Figure 6 displays the datasets. Some observations are highlighted in black.

The electricity demand curves are quite overlapping, but there are some notable differences in shape that distinguish the business day observations from the weekend ones. Typically, the business days have a much shorter daily low demand and a sharper rise in the level of demand towards the office hours compared to the two weekend days. The latter have a lower daily mean, that is also typically reached one to two hours later.

4.2. Classification results

We first illustrate the benefits of using different values of k in FD_j^k . To do so, maximum depth classification was conducted on the Australian weather dataset, for both groupings and for $J = 1, 2, 3$ and $k \in \{k' \in \mathbb{Z}: 1 \leq |k'| \leq 20\} \cup \{\pm\infty\}$. Figure 7 provides plots of the leave-one-out misclassification rates as a function of k , for each value of J .

As can be seen, for the North–South problem, FD_j^k performs consistently well with each J on low values of k , and maintains good performance until $k \approx 0$ for both $J = 2$ and $J = 3$. In the West–East context, there is a clear local minimum in the misclassification rate with a different value of k for each J . Contrarily to the North–South case, the values of k that minimize the misclassification rate are associated with low moments. Therefore, depending on the case, a depth emphasizing either centrality or extremality should be utilized. This highlights the fact that FD_j^k allows to adaptively select an appropriate depth for classification through the cross-validation on k . A similar exercise was conducted for the Adelaide dataset and guided the choices $J = 2$ and $k = -10$ in the motivating example in Section 2.

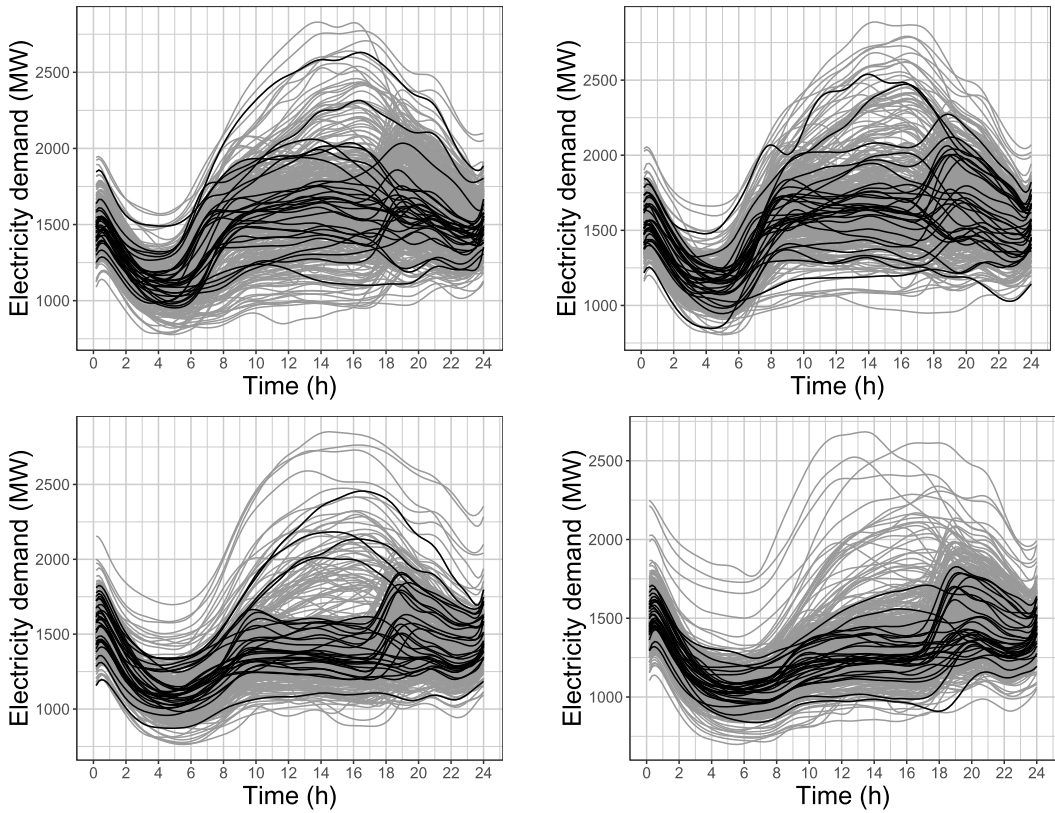


Figure 6. Adelaide electricity demand curves over Mondays (top left), Tuesdays (top right), Saturdays (bottom left) and Sundays (bottom right) between years 1997 and 2007, with randomly chosen observations highlighted in black.

Table 1 gives the leave-one-out misclassification rates for all supervised classification problems and all classification procedures described above. As explained above, all parameters k were chosen using leave-one-out cross-validation. Remarkably, FD_j^k performs consistently well for all problems.

Table 1. Leave-one-out misclassification rates for the classification problems described above, based on maximum depth classification with FD_j^k , $MFHD$, and $KFSD$, and based on the PC , PLS and kNN classifiers

	FD_1^k	FD_2^k	FD_3^k	$MFHD$	$KFSD$	PC	PLS	kNN
North–South	0.100	0.084	0.068	0.258	0.316	0.084	0.136	0.047
West–East	0.074	0.100	0.084	0.100	0.121	0.084	0.168	0.037
Tecator	0.316	0.112	0.093	0.216	0.326	0.093	0.037	0.191
Monday–Saturday	0.211	0.123	0.093	0.201	0.221	0.096	0.245	0.069
Monday–Sunday	0.176	0.103	0.098	0.156	0.158	0.055	0.172	0.060
Tuesday–Saturday	0.189	0.103	0.077	0.173	0.188	0.080	0.222	0.052
Tuesday–Sunday	0.122	0.058	0.044	0.130	0.121	0.012	0.136	0.023

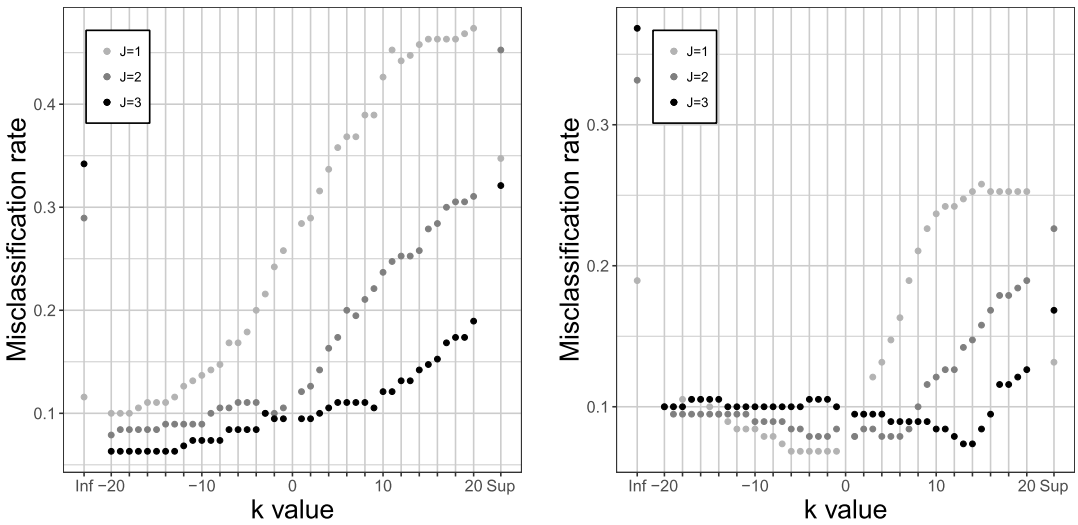


Figure 7. Leave-one-out misclassification rates of FD_J^k as a function of k for the North–South (left) and West–East (right) groupings of the Australian dataset, for $J = 1, 2, 3$.

First, FD_2^k and FD_3^k surpass other depth-based methods in all the cases considered. These two classifiers also outperform FD_1^k except for the West–East grouping of the Australian dataset. In most cases, except for electricity consumption classification on Monday–Sunday and Tuesday–Sunday, the PC classifier is slightly surpassed by FD_J^k for at least one value of J . This is all the more so interesting as PC is an optimal classifier under certain assumptions on the generating process (Delaigle and Hall [10]). PLS is the best classifier for the Tecator dataset, but it is consistently poor in all other examples. Finally, kNN has good misclassification rates in most cases, but it is outperformed by a large margin in the Tecator dataset.

Computationally, the complexity of FD_J^k is comparable to that of the J -dimensional halfspace depth, which in turn has a high computational complexity as a function of J and n . As such, efficient computation of halfspace depth has been a topic of rigorous research in the literature. There have been many contributions of both approximate and exact algorithms for fast computation of halfspace depth. See, for example, Hallin, Paindaveine and Šiman [21], Liu, Mosler and Mozharovskiy [28], Pokotylo, Mozharovskiy and Dyckerhoff [42] and Shao and Zuo [45] for further details. Out of the considered classifiers, kNN was clearly the fastest, followed by FD_1^k and $KFSD$. The PC and PLS classifiers, as well as the depth methods $MFHD$ and FD_2^k had similar computing times but were significantly slower than the previous methods. Out of the tested methods, FD_3^k was the slowest. A table with approximate average computation times of all the considered procedures can be found in the on-line supplemental document (Nagy et al. [40]).

5. Discussion and future prospects

In this paper, we introduced an adaptive class of functional depths. Their asymptotic properties were amply detailed and their good finite-sample performances were illustrated on several data examples. In this final section, we discuss two extra topics that relate to the proposed depths. First, we discuss modifications that can improve the classification procedures described in Section 4. Finally, we develop a bootstrap test for functional (halfspace) symmetry center.

Table 2. Leave-one-out misclassification rates for the classification problems described in Section 4, based on maximum depth classification with wFD_j^k

	North–South	West–East	Tec	Mon–Sat	Mon–Sun	Tue–Sat	Tue–Sun
wFD_2^k	0.063	0.058	0.065	0.094	0.084	0.055	0.031
wFD_3^k	0.058	0.063	0.037	0.077	0.068	0.035	0.025

5.0.0.4. *Improving on the classification procedure.* Section 4 illustrates the good classification performances of procedures built on the adaptive depth proposed in this paper. While we did not explore the theoretical consistency of the associated classifiers (this is topic for further research), we now discuss two practical modifications one could implement in order to improve classification.

First, one could consider alternatives to maximum depth classification by using best separation rules in the depth-depth plot, as was done in Li, Cuesta-Albertos and Liu [25]. We conducted an exploratory simulation study which showed that the small gain in misclassification rates was negligible compared to the increase in computational burden required from the second layer of cross-validation on the best separating line.

A more significant gain was obtained through a weighted approach, that is, by using the weighted depth (in the sequel, $w : [0, 1]^J \rightarrow [0, \infty)$ denotes a weight function)

$$wFD_j^k(x; w, P) = \left(\int_{[0,1]^J} w(t) (D(x(t); P_t) + 1/2)^k dt \right)^{1/k} - 1/2$$

in the maximum depth classification procedure. Since the values of the halfspace depths in each group might differ very little on some regions of $[0, 1]^J$, it is natural to select a weight function that discards such regions. Table 2 displays the misclassification rates obtained by assigning zero weight to the region with small halfspace differences. To be more precise, let $E_P[D(X(t); Q)]$ denote the expectation under $X \sim P \in \mathcal{P}(L^2([0, 1]))$ of the halfspace depth of $X(t)$ under $Q \in \mathcal{P}(\mathbb{R}^J)$ and, for distributions P_0 and P_1 , set $S_0(\lambda) = \{t : E_{P_0}[D(X(t); P_{0,t}) - D(X(t); P_{1,t})] > \lambda\}$, $S_1(\lambda) = \{t : E_{P_1}[D(X(t); P_{1,t}) - D(X(t); P_{0,t})] > \lambda\}$, and $w(t) = \mathbb{I}[t \in S_0(\lambda) \cup S_1(\lambda)]$. In practice, λ was chosen so that $S_0(\lambda) \cup S_1(\lambda)$ spans 25% of the area of $[0, 1]^J$. The misclassification rates show an improvement on the rates presented in Table 1. Interestingly, the procedure does not require an extra cross-validation step as the sets $S_0(\lambda)$ and $S_1(\lambda)$ can be constructed during the leave-one-out scheme on the sample elements.

In addition to the leave-one-out classification, the X -fold out-of-sample classification scheme described in Section 4 was performed for each dataset. In the training set, the weight function was first determined as described above, and the parameter k was tuned to minimize the leave-one-out misclassification rate with the chosen weight function. Then, the weight function and the corresponding k were used in leave-one-out classification of the testing set. The average out-of-sample misclassification rates and standard deviations (in brackets) are presented in Table 3. Aside from a few exceptions, the misclassification rates are similar to the rates presented in Table 2.

Figure 8 displays a heatmap of the mean differences for the low fat content observations of the Tecator dataset for $J = 2$. This plot confirms (see Figure 5) that the region with starkest discrimination is obtained for $t_1, t_2 \in [900, 950]$.

5.0.0.5. *Testing for symmetry.* The distributional asymptotics devised in Theorem 3 open up possibilities for inference. As an example, we propose to test whether a given function $x \in L^2([0, 1])$ is a center of J -symmetry of $P \in \mathcal{P}(L^2([0, 1]))$. By that, we mean that each marginal distribution $P_t \in \mathcal{P}(\mathbb{R}^J)$

Table 3. Average (over all test samples) out-of-sample misclassification rates (their standard deviations are provided in brackets) for the classification problems described in Section 4, based on maximum depth classification with wFD_J^k

	North–South	West–East	Tecator	Mon–Sat	Mon–Sun	Tue–Sat	Tue–Sun
wFD_2^k	0.072 (0.016)	0.089 (0.016)	0.063 (0.014)	0.141 (0.020)	0.091 (0.0003)	0.065 (0.014)	0.029 (0.0002)
wFD_3^k	0.066 (0.007)	0.086 (0.021)	0.061 (0.026)	0.099 (0.023)	0.085 (0.0009)	0.047 (0.008)	0.024 (0.0003)

of x satisfies $P_t(\{x(t)\}) = 0$ and is (halfspace) symmetric about $x(t)$, that is, $D(x(t); P_t) = 1/2$ for all $t \in [0, 1]^J$ (Zuo and Serfling [48]). Since x is then a maximizer of FD_J^k for any $|k| > 1$, with $FD_J^k(x; P) = 1/2$, this suggests to use the statistic $T_n^k = \sqrt{n}(1/2 - FD_J^k(x; P_n))$ to test the null hypothesis H_0 that x is a center of J -symmetry of P , against general alternatives. Theorem 3 asserts that the random quantity T_n^k converges in distribution to a non-degenerate positive random variable. Its distribution can be expressed as an integral of a properly weighted infimum of Gaussian processes. The asymptotic distribution of T_n^k under H_0 depends on P and on k . We propose to use nonparametric bootstrap to approximate it.

For $J = 1$, the procedure reduces to a test about the cross-sectional median, the deepest curve with respect to $FD_1^k(\cdot; P)$. In this situation, the bootstrap replicates are straightforward to generate under H_0 . Denote by x_m the cross-sectional median of P_n (or a representative if the median is not unique). By definition, $FD_1^k(x_m; P_n) = 1/2$, and x_m plays the role of the true center of symmetry for P_n and for the bootstrap replicates. For $b = 1, \dots, B$ with B a large integer, we propose to generate $T_{n,b}^{k,*} = \sqrt{n}(1/2 - FD_1^k(x_m; P_{n,b}^*))$, where for each b we write $P_{n,b}^*$ for the empirical measure that corresponds

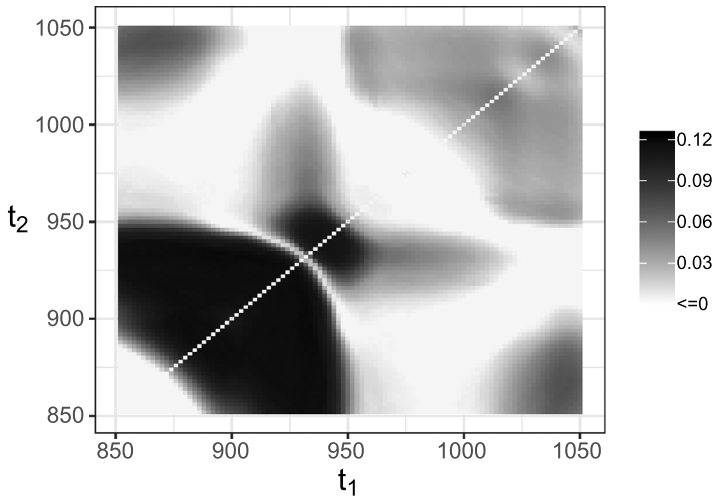


Figure 8. Average values of $D(x(t); P_{low,t}) - D(x(t); P_{high,t})$, for $t = (t_1, t_2)^T$, over all observations x with low fat content in the Tecator dataset. Here, P_{low} (resp. P_{high}) refers to the empirical distribution associated with the low (resp. high) fat content group.

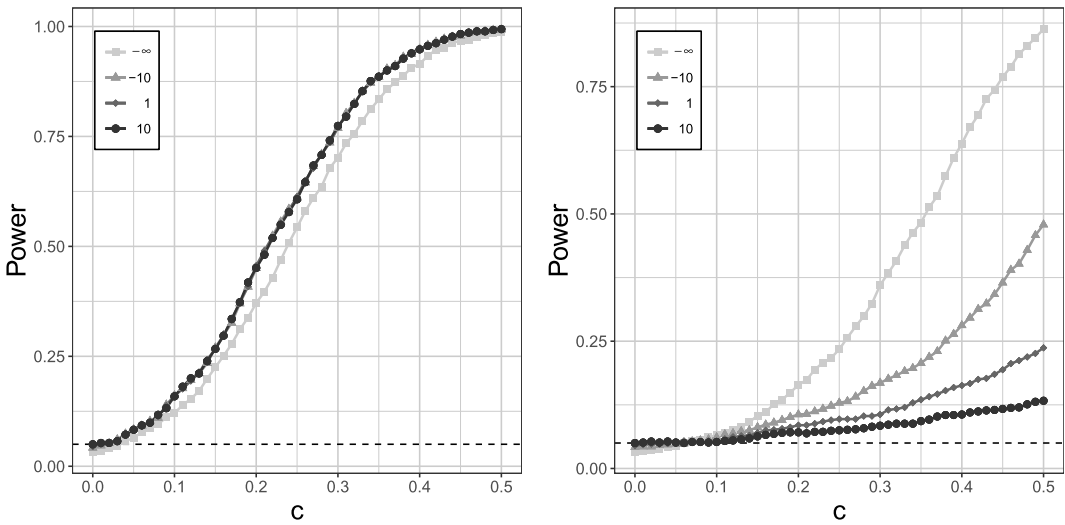


Figure 9. Simulated power functions for symmetry center testing based on T_n^k with $k \in \{-\infty, -10, 1, 10\}$ under location shifts (left) and peak modifications (right) of the true center of symmetry. The size of the tests is set to 0.05 (horizontal line). Bootstrap distributions were simulated with $B = 100$ and powers were approximated using 1000 independent runs. On the left panel, the curves for $k \neq -\infty$ overlap.

to an independent random selection of n functions from the original sample P_n with replacement. The p-value of our test is given by the usual expression $(1 + \sum_{b=1}^B \mathbb{I}[T_{n,b}^{k,*} > T_n^k]) / (1 + B)$.

We performed a short simulation study to assess the performance of the proposed tests. We generated random samples of size $n = 100$ from a centered Gaussian process with covariance function $E[X(t_1)X(t_2)] = \exp(-|t_1 - t_2|/3)$, $t_1, t_2 \in [0, 1]$, and tested whether its center of 1-symmetry is a constant function c (location shift scenario) or a function $c(1 - |4t - 2|)_+$ for $(x)_+$ the positive part of x (peak scenario), $c \in [0, 1/2]$. Case $c = 0$ corresponds to a situation under H_0 , while for $c > 0$, the alternative is true. The power functions presented in Figure 9 show that the tests perform quite well, in agreement with the results from Section 4. Parameter k allows to fine tune the testing procedure to emphasize different traits of the functional data – choice $k = 1$ rewards the average cross-sectional depth and performs well against alternatives of vertical shifts, while low values of k increase sensitivity against peaked alternatives. For $J > 1$, it is not obvious how to generate bootstrap resamples under H_0 . Theoretical validation of the proposed bootstrap procedure is out of the scope of the current paper, and will be subject of future research.

Appendix

This appendix collects all proofs of the theoretical results from the paper.

A.1. Proof of Theorem 1

We start by noting that by the boundedness of the domain $[0, 1]^J$ and measurability of every function involved (Nagy [36]), all integrals are well defined and finite.

Recall that for $x : [0, 1] \rightarrow \mathbb{R}$ and $\mathbf{t} = (t_1, \dots, t_J)^\top \in [0, 1]^J$ we denote $x(\mathbf{t}) = (x(t_1), \dots, x(t_J))^\top \in \mathbb{R}^J$, and $P_{\mathbf{t}} \equiv P_{(t_1, \dots, t_J)^\top} \in \mathcal{P}(\mathbb{R}^J)$ stands for the marginal distribution of $X(\mathbf{t}) \equiv (X(t_1), \dots, X(t_J))^\top$ with $X \sim P$. Similarly, let $P_{n,\mathbf{t}} \in \mathcal{P}(\mathbb{R}^J)$ stand for the marginal distribution of the empirical measure $P_n \in \mathcal{P}(L^2([0, 1]))$ at \mathbf{t} . With this notation, define functions

$$d : [0, 1]^J \rightarrow [1/2, 3/2] : \mathbf{t} \mapsto D(x(\mathbf{t}); P_{\mathbf{t}}) + 1/2,$$

$$d_n : [0, 1]^J \rightarrow [1/2, 3/2] : \mathbf{t} \mapsto D(x(\mathbf{t}); P_{n,\mathbf{t}}) + 1/2.$$

Universal almost sure consistency

Let us start with $k \geq 1$. In this case, one can write for any $x \in L^2([0, 1])$

$$\begin{aligned} |FD_J^k(x; P_n) - FD_J^k(x; P)| &= \left| \|d_n\|_k - \|d\|_k \right| \leq \|d_n - d\|_k \\ &= \left(\int_{[0,1]^J} |d_n(\mathbf{t}) - d(\mathbf{t})|^k \, d\mathbf{t} \right)^{1/k}, \end{aligned} \tag{A.1}$$

where we used that $\|\cdot\|_k$ is a norm. From this it follows that

$$\sup_{x \in L^2([0,1])} |FD_J^k(x; P_n) - FD_J^k(x; P)| \leq \left(\int_{[0,1]^J} \sup_{x \in L^2([0,1])} |d_n(\mathbf{t}) - d(\mathbf{t})|^k \, d\mathbf{t} \right)^{1/k},$$

and the last expression vanishes almost surely with $n \rightarrow \infty$ using the same technique as in the proof of Nagy et al. [39], Theorem 5.3, see also Nagy and Ferraty [37], Theorem 2, for its extension to the $L^2([0, 1])$ -space.

Now, let $k = -l \leq -1$. Because for $a, b > \lambda > 0$ we have

$$\left| \frac{1}{a} - \frac{1}{b} \right| \leq \frac{1}{\lambda^2} |a - b|,$$

by (3.2) we can write

$$\begin{aligned} |FD_J^k(x; P_n) - FD_J^k(x; P)| &= \left| \frac{1}{\|1/d_n\|_l} - \frac{1}{\|1/d\|_l} \right| \leq \frac{9}{4} \left| \|1/d_n\|_l - \|1/d\|_l \right| \\ &\leq \frac{9}{4} \|1/d_n - 1/d\|_l = \frac{9}{4} \left(\int_{[0,1]^J} \left| \frac{1}{d_n(\mathbf{t})} - \frac{1}{d(\mathbf{t})} \right|^l \, d\mathbf{t} \right)^{1/l} \\ &\leq 9 \left(\int_{[0,1]^J} |d_n(\mathbf{t}) - d(\mathbf{t})|^l \, d\mathbf{t} \right)^{1/l}. \end{aligned} \tag{A.2}$$

Since we bounded the expression for any $x \in L^2([0, 1])$ in the same way as in (A.1), we can continue as in the proof for positive k , and conclude that also in this setting the universal almost sure convergence holds true as desired.

\sqrt{n} -universal consistency

Starting from (A.1) and (A.2) for cases $k \geq 1$ and $k \leq -1$, respectively, it is enough to show that for all $k \geq 1$

$$\int_{[0,1]^J} \sup_{x \in L^2([0,1])} |d_n(\mathbf{t}) - d(\mathbf{t})|^k \, d\mathbf{t} = \mathcal{O}_P(n^{-k/2}),$$

or, equivalently, that

$$\int_{[0,1]^J} \sup_{x \in L^2([0,1])} (\sqrt{n} |d_n(\mathbf{t}) - d(\mathbf{t})|)^k d\mathbf{t} = \mathcal{O}_P(1).$$

Let \mathcal{H} be the collection of all halfspaces in \mathbb{R}^J . Since (Donoho and Gasko [11], formula (6.6))

$$\sup_{x \in L^2([0,1])} |d_n(\mathbf{t}) - d(\mathbf{t})| \leq \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)|, \tag{A.3}$$

it is enough to show that for all $k \geq 1$

$$\int_{[0,1]^J} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| \right)^k d\mathbf{t} = \mathcal{O}_P(1), \tag{A.4}$$

which can be read as the uniform tightness of the sequence of measures on the left-hand side of (A.4).

By the Vapnik-Červonenkis property of \mathcal{H} , an extension of the Dvoretzky–Kiefer–Wolfowitz inequality (Massart [32]) that can be found in Dudley [12], Section 6.5, gives that, for any $\varepsilon > 0$, there exists $K = K(\varepsilon)$ such that for all $v > 0$

$$\sup_{\mathbf{t} \in [0,1]^J} \mathbb{P} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| > v \right) \leq K \exp(-(2 - \varepsilon)v^2). \tag{A.5}$$

This inequality provides a bound on the moments of the supremum of the empirical halfspace process

$$\begin{aligned} \mathbb{E} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| \right)^{2k} &= \int_0^\infty \mathbb{P} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| > v^{1/(2k)} \right) dv \\ &\leq \int_0^\infty K \exp(-(2 - \varepsilon)v^{1/k}) dv, \end{aligned}$$

where the right-hand side is finite, and does not depend on either n or \mathbf{t} . We obtain that

$$\sup_{\mathbf{t} \in [0,1]^J} \sup_{n=1,2,\dots} \mathbb{E} \left(\left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| \right)^k \right)^2 = M < \infty.$$

Finally, with the help of Markov’s inequality and Fubini’s theorem we get that for all $L > 0$

$$\begin{aligned}
 & \sup_{n=1,2,\dots} \mathbb{P} \left(\left| \int_{[0,1]^J} (\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)|)^k dt \right| > L \right) \\
 & \leq \sup_{n=1,2,\dots} \frac{\mathbb{E} \left| \int_{[0,1]^J} (\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)|)^k dt \right|}{L} \\
 & \leq \sup_{n=1,2,\dots} \frac{\mathbb{E} \int_{[0,1]^J} |\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)||^k dt}{L} \\
 & = \sup_{n=1,2,\dots} \frac{\int_{[0,1]^J} \mathbb{E} |\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)||^k dt}{L} \\
 & \leq \frac{\sup_{t \in [0,1]^J} \sup_{n=1,2,\dots} \mathbb{E} (\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)|)^k}{L} \\
 & \leq \frac{\sqrt{M}}{L},
 \end{aligned}$$

and the desired (A.4) is verified. This completes the proof.

A.2. Proof of Theorem 2

Recall first the uniform law of the iterated logarithm for halfspaces in Euclidean spaces (Kuelbs and Dudley [22], Corollary 2.4). By that law it holds true that for any $Q \in \mathcal{P}(\mathbb{R}^J)$,

$$\limsup_{n \rightarrow \infty} \sup_{H \in \mathcal{H}} |Q_n(H) - Q(H)| \sqrt{\frac{2n}{\log \log n}} = 1 \quad \text{a.s.},$$

with Q_n the empirical measure of a random sample of size n from Q . The rates of convergence (2) for all values of $|k| \geq 1$ then follow from this formula, and

$$\begin{aligned}
 & \limsup_{n \rightarrow \infty} \sup_{x \in L^2([0,1])} |FD_J^k(x; P_n) - FD_J^k(x; P)| \\
 & \leq C_k \limsup_{n \rightarrow \infty} \left(\int_{[0,1]^J} \sup_{x \in L^2([0,1])} |d_n(t) - d(t)|^{|k|} dt \right)^{1/|k|} \\
 & \leq C_k \limsup_{n \rightarrow \infty} \left(\int_{[0,1]^J} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)|^{|k|} dt \right)^{1/|k|} \\
 & = C_k \left(\limsup_{n \rightarrow \infty} \int_{[0,1]^J} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)|^{|k|} dt \right)^{1/|k|} \\
 & \leq C_k \left(\int_{[0,1]^J} \left(\limsup_{n \rightarrow \infty} \sup_{H \in \mathcal{H}} |P_{n,t}(H) - P_t(H)| \right)^{|k|} dt \right)^{1/|k|}.
 \end{aligned}$$

The chain of three inequalities above holds true with $C_k = 1$ for $k \geq 1$ and $C_k = 9$ for $k \leq -1$, and follows from bounds (A.1) and (A.2), respectively, (A.3), and the reverse Fatou lemma (Dudley [13],

Lemma 4.3.3), in this order. The final almost sure result for the integral is obtained in the same spirit as the strong consistency in Nagy et al. [39], Theorem 5.3.

A.3. Proof of Theorem 3

In the first part of the proof we establish the result for $k = 1$. We use a result of Grinblat [19], Theorem 3, stated for $p = 1$.

Theorem 5 (Grinblat [19]). *Let ξ_n be a sequence of random processes with the following properties:*

(G₀) $\xi_n : [0, 1]^J \times \Omega \rightarrow \mathbb{R}$ is jointly measurable with respect to the product of the Lebesgue measure on $[0, 1]^J$, and \mathbb{P} ;

(G₁) for any $t_1, \dots, t_d \in [0, 1]^J$ the sequence $\{(\xi_n(t_1), \dots, \xi_n(t_d))^T\}_{n=1}^\infty$ converges in distribution to $(\xi(t_1), \dots, \xi(t_d))^T$;

(G₂) $\sup_{t \in [0, 1]^J} \sup_{n=1, 2, \dots} E |\xi_n(t)| < \infty$;

(G₃) $E |\xi_n(t)| \xrightarrow{n \rightarrow \infty} E |\xi(t)|$ for all $t \in [0, 1]^J$.

Then for any continuous functional f on $L^1([0, 1]^J)$ the weak convergence $f(\xi_n) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} f(\xi)$ holds true.

We apply Theorem 5 to the mapping

$$f : L^1([0, 1]^J) \rightarrow \mathbb{R} : \varphi \mapsto \int_{[0, 1]^J} \varphi(t) dt$$

and the process defined for $t \in [0, 1]^J$ and $\omega \in \Omega$ as

$$\xi_n(t) = \xi_n(t, \omega) = \sqrt{n} (D(x(t); P_{n,t}(\omega)) - D(x(t); P_t)).$$

The dependence of the process ξ_n on the random element ω will be suppressed in notation. We shall verify conditions (G₀)–(G₃) from Theorem 5.

Condition (G₀) is a corollary of Nagy [36], Theorem 1. Already for $k = 1$, the proof of (G₁) is rather long and technical. Therefore, the rest of the proof is split into six parts. The first four of them are devoted to (G₁). In part V below we establish

(G₄) $\sup_{t \in [0, 1]^J} \sup_{n=1, 2, \dots} E |\xi_n(t)|^2 < \infty$,

and show that this condition together with (G₁) is sufficient for (G₂) and (G₃). In the final part VI of this proof, the result of Theorem 3 for general k is obtained.

Let $\mathcal{H}_* = \mathcal{H} \cup \mathbb{R}^J$ stand for the collection of generalized halfspaces – all halfspaces in \mathbb{R}^J , and \mathbb{R}^J . For $d = 1, 2, \dots$ denote by

$$\mathcal{H}_*^d = \{H = H_1 \times \dots \times H_d : H_j \in \mathcal{H}_* \text{ for all } j = 1, \dots, d\}$$

the collection of all Cartesian products of d generalized halfspaces. Note that formally \mathcal{H}_*^d must be distinguished from the collection of all d -tuples of generalized halfspaces denoted by $(\mathcal{H}_*)^d$. Nevertheless, the canonical bijection $(\mathcal{H}_*)^d \rightarrow \mathcal{H}_*^d : (H_1, \dots, H_d) \mapsto H_1 \times \dots \times H_d$ makes the two spaces equivalent.

Part I: Joint empirical process

We begin with some auxiliary derivations. Fix $d = 1, 2, \dots$ and $t_1, \dots, t_d \in [0, 1]^J$, and for $X \sim P \in \mathcal{P}(L^2([0, 1]))$ denote by $P_{(t_1, \dots, t_d)} \in \mathcal{P}(\mathbb{R}^{Jd})$ the joint distribution of the (Jd) -dimensional random vector $(X(t_1)^\top, \dots, X(t_d)^\top)^\top \in \mathbb{R}^{Jd}$. Its marginal distributions are $P_{t_1}, \dots, P_{t_d} \in \mathcal{P}(\mathbb{R}^J)$. Similarly, write $P_{n,(t_1, \dots, t_d)} \in \mathcal{P}(\mathbb{R}^{Jd})$ for the corresponding empirical distribution based on a random sample of size n from P .

Consider the sequence of processes

$$\sqrt{n}(P_{n,(t_1, \dots, t_d)}(\mathbf{H}) - P_{(t_1, \dots, t_d)}(\mathbf{H})) \tag{A.6}$$

indexed by $\mathbf{H} \in \mathcal{H}_*^d$. Since the collection of halfspaces \mathcal{H} in \mathbb{R}^J has the Vapnik-Červonenkis property, also collections \mathcal{H}_* and \mathcal{H}_*^d are Vapnik-Červonenkis classes of sets in \mathbb{R}^J and \mathbb{R}^{Jd} , respectively (van der Vaart and Wellner [47], Lemma 2.6.17). Therefore, the process (A.6) is Donsker (van der Vaart and Wellner [47], Sections 2.5 and 2.6.1), and weakly convergent in $\ell^\infty(\mathcal{H}_*^d)$ (to a $P_{(t_1, \dots, t_d)}$ -Brownian bridge) as desired.

Part II: Marginal empirical processes

The following lemma enables to transfer the weak convergence result from a joint univariate empirical process to a vector of its marginals.

Lemma 1. *For a set T such that $s \in T$ and $d = 1, 2, \dots$, let X_n converge weakly to X in $\ell^\infty(T^d)$. Define*

$$\pi_j : T \rightarrow T^d : t_j \mapsto (\underbrace{s, \dots, s}_{(j-1)\text{-times}}, t_j, \underbrace{s, \dots, s}_{(d-j)\text{-times}})^\top.$$

Then $\{(X_n \circ \pi_1, \dots, X_n \circ \pi_d)\}_{n=1}^\infty$ converges weakly to $(X \circ \pi_1, \dots, X \circ \pi_d)^\top$ in the space $\ell^\infty(T)^d$.

Proof. For $x, y \in \ell^\infty(T^d)$ and $h : \ell^\infty(T^d) \rightarrow (\ell^\infty(T))^d : x \mapsto (x \circ \pi_1, \dots, x \circ \pi_d)^\top$ we have

$$\|h(x) - h(y)\| = \sum_{j=1}^d \|x \circ \pi_j - y \circ \pi_j\| \leq d\|x - y\|,$$

which means that h is Lipschitz continuous. The result follows by the continuous mapping theorem (Dudley [13], Theorem 9.3.7). □

By Lemma 1 with $T = \mathcal{H}_*$, X_n the process in (A.6), $s = \mathbb{R}^J$, and $t_j = H_j \in \mathcal{H}$, we obtain that with

$$\begin{aligned} X_n \circ \pi_j(H_j) &= \sqrt{n}(P_{n,(t_1, \dots, t_d)}(\mathbb{R}^J \times \dots \times \mathbb{R}^J \times H_j \times \mathbb{R}^J \times \dots \times \mathbb{R}^J) \\ &\quad - P_{(t_1, \dots, t_d)}(\mathbb{R}^J \times \dots \times \mathbb{R}^J \times H_j \times \mathbb{R}^J \times \dots \times \mathbb{R}^J)) \\ &= \sqrt{n}(P_{n,t_j}(H_j) - P_{t_j}(H_j)), \end{aligned}$$

the sequence of processes

$$\sqrt{n}(P_{n,t_1}(H_1) - P_{t_1}(H_1), \dots, P_{n,t_d}(H_d) - P_{t_d}(H_d))^\top,$$

indexed by $(H_1, \dots, H_d)^\top \in (\mathcal{H}_*)^d$, converges weakly in $\ell^\infty(\mathcal{H}_*)^d$ to the d -dimensional image of the $P_{(t_1, \dots, t_d)}$ -Brownian bridge given by the transformation used in Lemma 1.

Part III: Approximation of the depth process

To verify **(G₁)**, we now combine the technical result from part II with the method used in the proof of Massé [33], Theorem 2.1. Following the notation from that paper, for $t \in [0, 1]^J$, let $V(t) \subset S^{J-1}$ denote the collection of all inner normals of minimal halfspaces of measure $P_t \in \mathcal{P}(\mathbb{R}^J)$ at $x(t)$ (recall that, throughout the proof, $x \in L^2([0, 1])$ is fixed). For each $t \in [0, 1]^J$, we assumed that $P_t \in \mathcal{P}(\mathbb{R}^J)$ satisfies **(S)**. Under this condition, Massé [33], Proposition 4.5, argues that the mapping $S^{J-1} \rightarrow [0, 1]: u \mapsto P_t(H[x(t), u])$ is continuous, in particular the set of its arguments of minima $V(t)$ is closed in S^{J-1} . Similarly as in the first paragraph of Massé [33], Section 5.2, for any t we cover the set $V(t)$ by isolating sets. By isolating sets we mean a system of closed subsets $\{U_t(v)\}_{v \in V(t)}$ of vectors in S^{J-1} , such that for each $v \in V(t)$

$$U_t(v) \cap V(t) = \{v\} \quad \text{and} \quad S^{J-1} = \bigcup_{v \in V(t)} U_t(v).$$

In contrast to Massé [33] who bases his proof on an arbitrary covering of S^{J-1} by isolating sets, in our proof it will be useful to introduce a specific covering given by

$$U_t(v) = \left\{ u \in S^{J-1} : \|u - v\| \leq \inf_{v' \in V(t)} \|u - v'\| \right\}. \tag{A.7}$$

Because distance to a set $V(t)$ is a continuous function, each set $U_t(v)$ is closed. At the same time, the collection $\{U_t(v) : v \in V(t)\}$ covers S^{J-1} because for each $u \in S^{J-1}$ there exists an element of the closed set $V(t)$ that minimizes the distance from u to points of $V(t)$. Finally, for $v, \tilde{v} \in V(t)$ such that $\tilde{v} \in U_t(v)$ but $\tilde{v} \neq v$ we necessarily have

$$0 \leq \|\tilde{v} - v\| \leq \inf_{v' \in V(t)} \|\tilde{v} - v'\| = \|\tilde{v} - \tilde{v}\| = 0,$$

which gives a contradiction, meaning that the sets (A.7) are indeed isolating.

For $j = 1, \dots, d$, let

$$\mathcal{H}_{\min}(t_j) = \bigcup_{v_j \in V(t_j)} H[x(t_j), v_j] \subset \mathcal{H}_*$$

be the collection of all minimal halfspaces at point $x(t_j)$.

For $j = 1, \dots, d$ and $H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)$ let $\lambda_{n,j}$ be a random process indexed by $\mathcal{H}_{\min}(t_j)$ given by

$$\begin{aligned} &\lambda_{n,j}(H[x(t_j), v_j]) \\ &= \begin{cases} \sqrt{n} \left(\inf_{u_j \in U_{t_j}(v_j)} P_{n,t_j}(H[x(t_j), u_j]) - P_{t_j}(H[x(t_j), v_j]) \right) & \text{if } D(x(t_j); P_{t_j}) > 0, \\ 0 & \text{if } D(x(t_j); P_{t_j}) = 0. \end{cases} \end{aligned}$$

In what follows, we shall use also $\tilde{u}_j \in S^{J-1}$ as a shorthand for any direction that satisfies

$$\inf_{u_j \in U_{t_j}(v_j)} P_{n,t_j}(H[x(t_j), u_j]) = P_{n,t_j}(H[x(t_j), \tilde{u}_j]).$$

Note that \tilde{u}_j depends on t_j, v_j, n , and $\omega \in \Omega$. For each j , let

$$v_{n,j}(H[x(t_j), v_j]) = \sqrt{n}(P_{n,t_j}(H[x(t_j), v_j]) - P_{t_j}(H[x(t_j), v_j])).$$

By the same argument as in the proof of Massé [33], Lemma 5.4, we have

$$v_{n,j}(H[x(t_j), \tilde{u}_j]) \leq \lambda_{n,j}(H[x(t_j), v_j]) \leq v_{n,j}(H[x(t_j), v_j]).$$

For the vectors of processes $\lambda_n = (\lambda_{n,1}, \dots, \lambda_{n,d})^\top$ and $\mathbf{v}_n = (v_{n,1}, \dots, v_{n,d})^\top$ in the Cartesian product of ℓ^∞ spaces $\prod_{j=1}^d \ell^\infty(\mathcal{H}_{\min}(t_j))$ equipped with the norm given by the sum of the norms in its components, the above inequalities mean that

$$\begin{aligned} \|\lambda_n - \mathbf{v}_n\| &= \sum_{j=1}^d \|\lambda_{n,j} - v_{n,j}\| \\ &= \sum_{j=1}^d \sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - \lambda_{n,j}(H[x(t_j), v_j])) \\ &\leq \sum_{j=1}^d \sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])). \end{aligned} \tag{A.8}$$

Now we show that the bound on the right hand side of the previous display vanishes in outer probability uniformly in $v_j \in V(t_j)$ for all $j = 1, \dots, d$. To see this, write for $\varepsilon > 0$

$$\begin{aligned} &\mathbf{P}^* \left(\sum_{j=1}^d \sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])) > \varepsilon \right) \\ &\leq \mathbf{P}^* \left(\bigcup_{j=1}^d \left[\sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])) > \varepsilon/d \right] \right) \\ &\leq \sum_{j=1}^d \mathbf{P}^* \left(\sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])) > \varepsilon/d \right), \end{aligned}$$

where \mathbf{P}^* stands for the outer probability on Ω . By Fatou’s lemma,

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \mathbf{P}^* \left(\sum_{j=1}^d \sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])) > \varepsilon \right) \\ &\leq \sum_{j=1}^d \limsup_{n \rightarrow \infty} \mathbf{P}^* \left(\sup_{H[x(t_j), v_j] \in \mathcal{H}_{\min}(t_j)} (v_{n,j}(H[x(t_j), v_j]) - v_{n,j}(H[x(t_j), \tilde{u}_j])) > \varepsilon/d \right). \end{aligned}$$

Now, we use the derivations provided in the proof of Massé [33], Lemma 5.4, and apply them to each of the terms of the sum on the right-hand side above. From the final paragraph of the proof of Massé [33], Lemma 5.4, it then follows that the expression above vanishes if for each $t \in [0, 1]^J$, a sequence $\{v_n\}_{n=1}^\infty \subset V(t)$ such that $v_n \rightarrow v_0 \in V(t)$, and a sequence $\{u_n\}_{n=1}^\infty$ such that $u_n \in U_t(v_n)$ for each n and $u_n \rightarrow u_0 \in V(t)$, it necessarily holds true that $v_0 = u_0$. We demonstrate this by using the special form of the covering (A.7). Suppose for contradiction that $\Delta = \|u_0 - v_0\| > 0$. By convergence, for n large enough we know that $\|u_n - u_0\| \leq \Delta/4$ and at the same time $\|v_n - v_0\| \leq \Delta/4$. Because

$u_0 \in V(\mathbf{t})$ but also $u_n \in U_{\mathbf{t}}(v_n)$, the construction of the covering (A.7) gives that necessarily $\|u_n - v_n\| \leq \|u_n - u_0\| \leq \Delta/4$. This however means that

$$0 < \Delta = \|u_0 - v_0\| \leq \|u_0 - u_n\| + \|u_n - v_n\| + \|v_n - v_0\| \leq 3\Delta/4,$$

a contradiction giving that $u_0 = v_0$. This slight modification to the proof of Massé [33], Lemma 5.4, ensures that the vector process on the right-hand side of (A.8) vanishes in outer probability as $n \rightarrow \infty$, that is $\|\lambda_n - \mathbf{v}_n\| \xrightarrow[n \rightarrow \infty]{P^*} 0$. By van der Vaart and Wellner [47], Lemma 1.10.2, this implies that the two sequences of processes λ_n and \mathbf{v}_n have the same limit process; in particular, λ_n converges weakly in $\prod_{j=1}^d \ell^\infty(\mathcal{H}_{\min}(\mathbf{t}_j))$. Weak convergence of \mathbf{v}_n and its limit were established in part II of this proof.

Part IV: Continuous mapping theorem

By an argument similar to that in the proof of Massé [33], Lemma 5.5, the mapping

$$\begin{aligned} \mathcal{J} : \prod_{j=1}^d \ell^\infty(\mathcal{H}_{\min}(\mathbf{t}_j)) &\rightarrow \mathbb{R}^d \\ &: \boldsymbol{\phi} \mapsto \left(\inf_{v_1 \in V(\mathbf{t}_1)} \phi_1(H[x(\mathbf{t}_1), v_1]), \dots, \inf_{v_d \in V(\mathbf{t}_d)} \phi_d(H[x(\mathbf{t}_d), v_d]) \right)^\top \end{aligned}$$

for $\boldsymbol{\phi} = (\phi_1, \dots, \phi_d)^\top$ is continuous. To see this, let $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2 \in \prod_{j=1}^d \ell^\infty(\mathcal{H}_{\min}(\mathbf{t}_j))$ with $\boldsymbol{\phi}_i = (\phi_{i,1}, \dots, \phi_{i,d})^\top$ for $\phi_{i,j} \in \ell^\infty(\mathcal{H}_{\min}(\mathbf{t}_j))$ and $i = 1, 2, j = 1, \dots, d$. Then we can write

$$\begin{aligned} \|\mathcal{J}(\boldsymbol{\phi}_1) - \mathcal{J}(\boldsymbol{\phi}_2)\| &= \left(\sum_{j=1}^d \left| \inf_{v_j \in V(\mathbf{t}_j)} \phi_{1,j}(H[x(\mathbf{t}_j), v_j]) - \inf_{v_j \in V(\mathbf{t}_j)} \phi_{2,j}(H[x(\mathbf{t}_j), v_j]) \right|^2 \right)^{1/2} \\ &\leq \left(\sum_{j=1}^d \left(\sup_{v_j \in V(\mathbf{t}_j)} |\phi_{1,j}(H[x(\mathbf{t}_j), v_j]) - \phi_{2,j}(H[x(\mathbf{t}_j), v_j])| \right)^2 \right)^{1/2} \\ &\leq \sum_{j=1}^d \sup_{v_j \in V(\mathbf{t}_j)} |\phi_{1,j}(H[x(\mathbf{t}_j), v_j]) - \phi_{2,j}(H[x(\mathbf{t}_j), v_j])| \\ &= \|\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2\|. \end{aligned}$$

Therefore, we may apply \mathcal{J} to the vector process \mathbf{v}_n , and by the continuous mapping theorem (Dudley [12], Theorem 3.6.7) and part III of this proof we obtain that both $\mathcal{J}(\mathbf{v}_n)$, and $\mathcal{J}(\lambda_n) = (\xi_n(\mathbf{t}_1), \dots, \xi_n(\mathbf{t}_d))^\top$, converge weakly to $(\xi(\mathbf{t}_1), \dots, \xi(\mathbf{t}_d))^\top$, and Condition **(G₁)** is verified.

Part V: Condition (G₄)

Condition **(G₄)** is satisfied thanks to **(G₁)**, (A.3) and (A.5) from the proof of Theorem 1. Indeed,

$$\begin{aligned} \sup_{\mathbf{t} \in [0,1]^J} \sup_{n=1,2,\dots} \mathbb{E}|\xi_n(\mathbf{t})|^2 &\leq \sup_{\mathbf{t} \in [0,1]^J} \sup_{n=1,2,\dots} \mathbb{E} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| \right)^2 \\ &= \int_0^\infty \mathbb{P} \left(\sqrt{n} \sup_{H \in \mathcal{H}} |P_{n,\mathbf{t}}(H) - P_{\mathbf{t}}(H)| > v^{1/2} \right) dv \end{aligned}$$

$$\leq \int_0^\infty K \exp(-(2 - \varepsilon)v) dv,$$

where the last expression is finite, and independent of n and \mathbf{t} .

Now, condition (\mathbf{G}_2) follows from (\mathbf{G}_4) by direct application of Jensen’s inequality (Dudley [13], Theorem 10.2.6). Further, for $L > 0$ we have

$$\sup_{n=1,2,\dots} E(|\xi_n(\mathbf{t})| \mathbb{I}[|\xi_n(\mathbf{t})| > L]) \leq \sup_{n=1,2,\dots} E(|\xi_n(\mathbf{t})|^2 / L),$$

where the last expression vanishes as $L \rightarrow \infty$ by (\mathbf{G}_4) . Thus, $\{|\xi_n(\mathbf{t})|\}_{n=1}^\infty$ is uniformly integrable (Dudley [13], Section 10.3), and its convergence in distribution from (\mathbf{G}_1) implies the convergence of its expectation (Billingsley [2], Theorem 25.12). That gives (\mathbf{G}_3) .

Part VI: General result

In the case of general k , we start from Conditions (\mathbf{G}_0) – (\mathbf{G}_4) for $k = 1$, and verify their versions for the process

$$\xi_n(\mathbf{t}) = \sqrt{n} \left((D(x(\mathbf{t}); P_{n,\mathbf{t}}) + 1/2)^k - (D(x(\mathbf{t}); P_{\mathbf{t}}) + 1/2)^k \right).$$

Condition (\mathbf{G}_0) follows immediately. (\mathbf{G}_1) can be recovered from the base case $k = 1$ and the delta-method (van der Vaart and Wellner [47], Section 3.9) applied to the function

$$g: \mathbb{R}^d \rightarrow \mathbb{R}^d: (s_1, \dots, s_d)^\top \mapsto ((s_1 + 1/2)^k, \dots, (s_d + 1/2)^k)^\top.$$

The Jacobian of g at $(D(x(\mathbf{t}_1); P_{t_1}), \dots, D(x(\mathbf{t}_d); P_{t_d}))^\top$ is a diagonal matrix with $k(D(x(\mathbf{t}_j); P_{t_j}) + 1/2)^{k-1}$ as its j -th diagonal element.

To show (\mathbf{G}_4) note that for any $a, b \in [0, 1]$ and $k \in \mathbb{R}$ we have

$$|(a + 1/2)^k - (b + 1/2)^k| \leq |k| \max \left\{ \frac{1}{2^{k-1}}, \left(\frac{3}{2} \right)^{k-1} \right\} |a - b|.$$

Thus,

$$|\xi_n(\mathbf{t})| \leq \sqrt{n} |k| \max \left\{ \frac{1}{2^{k-1}}, \left(\frac{3}{2} \right)^{k-1} \right\} |D(x(\mathbf{t}); P_{n,\mathbf{t}}) - D(x(\mathbf{t}); P_{\mathbf{t}})|$$

and the moment bound from part V is enough to get (\mathbf{G}_4) also for general k .

By Theorem 5 again, we obtain the weak convergence of the process

$$\sqrt{n} \left(\int_{[0,1]^J} (D(x(\mathbf{t}); P_{n,\mathbf{t}}) + 1/2)^k dt - \int_{[0,1]^J} (D(x(\mathbf{t}); P_{\mathbf{t}}) + 1/2)^k dt \right).$$

Now it is enough to apply the delta-method again, this time with the function

$$g: \mathbb{R} \rightarrow \mathbb{R}: s \mapsto s^{1/k} - 1/2,$$

and the weak convergence of the moment integrated depths holds true as desired.

Acknowledgements

The authors greatly appreciate the insightful comments of an Associate Editor and a referee, which led to distinct improvements in the paper. The authors would also like to acknowledge the computational resources provided by the Aalto University School of Science “Science-IT” project. S. Nagy wishes to thank the Czech Science Foundation (grant 19-16097Y) and Charles University (grant PRIMUS/17/SCI/3). G. Van Bever would like to thank the Belgian FNRS (grant Crédit de recherche C 60/5-CDR/OL). P. Ilmonen and L. Viitasaari wish to thank the Väisälä foundation for its support.

Supplementary Material

R source codes and a supplemental document (DOI: [10.3150/20-BEJ1254SUPP](https://doi.org/10.3150/20-BEJ1254SUPP); .zip). Complete R source codes that enable replication of the analyses performed in the paper, and a supplemental document (.pdf) containing (i) the omitted X -fold out-of-sample classification performance of FD_J^k , and (ii) a table of approximate computation times of the considered methods for the leave-one-out classification of the Tecator dataset.

References

- [1] Biau, G., Bunea, F. and Wegkamp, M.H. (2005). Functional classification in Hilbert spaces. *IEEE Trans. Inf. Theory* **51** 2163–2172. MR2235289 <https://doi.org/10.1109/TIT.2005.847705>
- [2] Billingsley, P. (1995). *Probability and Measure*, 3rd ed. *Wiley Series in Probability and Mathematical Statistics*. New York: Wiley. MR1324786
- [3] Borggaard, C. and Thodberg, H.H. (1992). Optimal minimal neural interpretation of spectra. *Anal. Chem.* **64** 545–551.
- [4] Cérou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat.* **10** 340–355. MR2247925 <https://doi.org/10.1051/ps:2006014>
- [5] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Ann. Statist.* **42** 1203–1231. MR3224286 <https://doi.org/10.1214/14-AOS1226>
- [6] Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2017). Monge–Kantorovich depth, quantiles, ranks and signs. *Ann. Statist.* **45** 223–256. MR3611491 <https://doi.org/10.1214/16-AOS1450>
- [7] Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.* **109** 411–423. MR3180573 <https://doi.org/10.1080/01621459.2013.856795>
- [8] Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.* **22** 481–496. MR2336349 <https://doi.org/10.1007/s00180-007-0053-0>
- [9] Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivariate Anal.* **100** 753–766. MR2478196 <https://doi.org/10.1016/j.jmva.2008.08.002>
- [10] Delaigle, A. and Hall, P. (2012). Achieving near perfect classification for functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 267–286. MR2899863 <https://doi.org/10.1111/j.1467-9868.2011.01003.x>
- [11] Donoho, D.L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827. MR1193313 <https://doi.org/10.1214/aos/1176348890>
- [12] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics* **63**. Cambridge: Cambridge Univ. Press. MR1720712 <https://doi.org/10.1017/CBO9780511665622>
- [13] Dudley, R.M. (2002). *Real Analysis and Probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge: Cambridge Univ. Press. MR1932358 <https://doi.org/10.1017/CBO9780511755347>
- [14] Ferraty, F. and Vieu, P. (2003). Curves discrimination: A nonparametric functional approach **44** 161–173. MR2020144 [https://doi.org/10.1016/S0167-9473\(03\)00032-X](https://doi.org/10.1016/S0167-9473(03)00032-X)

- [15] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. New York: Springer. MR2229687
- [16] Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *TEST* **10** 419–440. MR1881149 <https://doi.org/10.1007/BF02595706>
- [17] Ghosh, A.K. and Chaudhuri, P. (2005). On maximum depth and related classifiers. *Scand. J. Stat.* **32** 327–350. MR2188677 <https://doi.org/10.1111/j.1467-9469.2005.00423.x>
- [18] Gijbels, I. and Nagy, S. (2015). Consistency of non-integrated depths for functional data. *J. Multivariate Anal.* **140** 259–282. MR3372567 <https://doi.org/10.1016/j.jmva.2015.05.012>
- [19] Grinblat, L.Š. (1976). A limit theorem for measurable random processes and its applications. *Proc. Amer. Math. Soc.* **61** 371–376. MR0423450 <https://doi.org/10.2307/2041344>
- [20] Hall, P., Poskitt, D.S. and Presnell, B. (2001). A functional data-analytic approach to signal discrimination. *Technometrics* **43** 1–9. MR1847775 <https://doi.org/10.1198/00401700152404273>
- [21] Hallin, M., Paindaveine, D. and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth. *Ann. Statist.* **38** 635–669. MR2604670 <https://doi.org/10.1214/09-AOS723>
- [22] Kuelbs, J. and Dudley, R.M. (1980). Log log laws for empirical measures. *Ann. Probab.* **8** 405–418. MR0573282
- [23] Kuelbs, J. and Zinn, J. (2013). Concerns with functional depth. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 831–855. MR3125749
- [24] Li, B., Van Bever, G., Oja, H., Sabolova, R. and Critchley, F. (2019). Functional independent component analysis: An extension of fourth order blind identification. Technical Report, Univ. Namur.
- [25] Li, J., Cuesta-Albertos, J.A. and Liu, R.Y. (2012). DD -classifier: Nonparametric classification procedure based on DD -plot. *J. Amer. Statist. Assoc.* **107** 737–753. MR2980081 <https://doi.org/10.1080/01621459.2012.688462>
- [26] Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18** 405–414. MR1041400 <https://doi.org/10.1214/aos/1176347507>
- [27] Liu, R.Y., Parelius, J.M. and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Ann. Statist.* **27** 783–858. MR1724033 <https://doi.org/10.1214/aos/1018031260>
- [28] Liu, X., Mosler, K. and Mozharovskiy, P. (2019). Fast computation of Tukey trimmed regions and median in dimension $p > 2$. *J. Comput. Graph. Statist.* **28** 682–697. MR4007750 <https://doi.org/10.1080/10618600.2018.1546595>
- [29] López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.* **104** 718–734. MR2541590 <https://doi.org/10.1198/jasa.2009.0108>
- [30] López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.* **55** 1679–1695. MR2748671 <https://doi.org/10.1016/j.csda.2010.10.024>
- [31] Magnano, L., Boland, J.W. and Hyndman, R.J. (2008). Generation of synthetic sequences of half-hourly temperature. *Environmetrics* **19** 818–835. MR2654607 <https://doi.org/10.1002/env.905>
- [32] Massart, P. (1986). Rates of convergence in the central limit theorem for empirical processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **22** 381–423. MR0871904
- [33] Massé, J.-C. (2004). Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli* **10** 397–419. MR2061438 <https://doi.org/10.3150/bj/1089206404>
- [34] Mosler, K. (2013). Depth statistics. In *Robustness and Complex Data Structures* (C. Becker, R. Fried and S. Kuhnt, eds.) 17–34. Heidelberg: Springer. MR3135871 https://doi.org/10.1007/978-3-642-35494-6_2
- [35] Mosler, K. and Polyakova, Y. (2016). General notions of depth for functional data. Preprint. Available at [arXiv:1208.1981](https://arxiv.org/abs/1208.1981).
- [36] Nagy, S. (2017). Integrated depth for measurable functions and sets. *Statist. Probab. Lett.* **123** 165–170. MR3598634 <https://doi.org/10.1016/j.spl.2016.12.012>
- [37] Nagy, S. and Ferraty, F. (2019). Data depth for measurable noisy random functions. *J. Multivariate Anal.* **170** 95–114. MR3913030 <https://doi.org/10.1016/j.jmva.2018.11.003>
- [38] Nagy, S., Gijbels, I. and Hlubinka, D. (2017). Depth-based recognition of shape outlying functions. *J. Comput. Graph. Statist.* **26** 883–893. MR3765352 <https://doi.org/10.1080/10618600.2017.1336445>

- [39] Nagy, S., Gijbels, I., Omelka, M. and Hlubinka, D. (2016). Integrated depth for functional data: Statistical properties and consistency. *ESAIM Probab. Stat.* **20** 95–130. MR3528619 <https://doi.org/10.1051/ps/2016005>
- [40] Nagy, S., Helander, S., Van Bever, G., Viitasaari, L. and Ilmonen, P. (2020). Supplement to “Flexible integrated functional depths.” <https://doi.org/10.3150/20-BEJ1254SUPP>
- [41] Narisetty, N.N. and Nair, V.N. (2016). Extremal depth for functional data and applications. *J. Amer. Statist. Assoc.* **111** 1705–1714. MR3601729 <https://doi.org/10.1080/01621459.2015.1110033>
- [42] Pokotylo, O., Mozharovskyi, P. and Dyckerhoff, R. (2019). Depth and depth-based classification with R package ddalpha. *J. Stat. Softw.* **91** 1–46.
- [43] Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR2168993
- [44] Sguera, C., Galeano, P. and Lillo, R. (2014). Spatial depth-based classification for functional data. *TEST* **23** 725–750. MR3274472 <https://doi.org/10.1007/s11749-014-0379-1>
- [45] Shao, W. and Zuo, Y. (2020). Computing the halfspace depth with multiple try algorithm and simulated annealing algorithm. *Comput. Statist.* **35** 203–226.
- [46] Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974)*, Vol. 2 523–531. MR0426989
- [47] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. New York: Springer. MR1385671 <https://doi.org/10.1007/978-1-4757-2545-2>
- [48] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482. MR1790005 <https://doi.org/10.1214/aos/1016218226>

Received November 2019 and revised March 2020