

The Area of the Convex Hull of Sampled Curves: a Robust Functional Statistical Depth Measure

Guillaume Staerman, Pavlo Mozharovskyi, Stéphan Clémençon

▶ **To cite this version:**

Guillaume Staerman, Pavlo Mozharovskyi, Stéphan Clémençon. The Area of the Convex Hull of Sampled Curves: a Robust Functional Statistical Depth Measure. PMLR 108:570-579, 2020. hal-03132996

HAL Id: hal-03132996

https://telecom-paris.hal.science/hal-03132996

Submitted on 5 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Area of the Convex Hull of Sampled Curves: a Robust Functional Statistical Depth Measure

Guillaume Staerman* **Pavlo Mozharovskyi*** **Stephan Clémençon***
*LTCI, Télécom Paris, Institut Polytechnique de Paris

Abstract

With the ubiquity of sensors in the IoT era, statistical observations are becoming increasingly available in the form of massive (multivariate) time-series. Formulated as unsupervised anomaly detection tasks, an abundance of applications like aviation safety management, the health monitoring of complex infrastructures or fraud detection can now rely on such functional data, acquired and stored with an ever finer granularity. The concept of *statistical depth*, which reflects centrality of an arbitrary observation w.r.t. a statistical population may play a crucial role in this regard, anomalies corresponding to observations with 'small' depth. Supported by sound theoretical and computational developments in the recent decades, it has proven to be extremely useful, in particular in functional spaces. However, most approaches documented in the literature consist in evaluating independently the centrality of each point forming the time series and consequently exhibit a certain insensitivity to possible shape changes. In this paper, we propose a novel notion of functional depth based on the area of the convex hull of sampled curves, capturing gradual departures from centrality, even beyond the envelope of the data, in a natural fashion. We discuss practical relevance of commonly imposed axioms on functional depths and investigate which of them are satisfied by the notion of depth we promote here. Estimation and computational issues are also addressed and various numerical experiments provide empirical evidence of the relevance of the approach proposed.

Proceedings of the 23[rd]International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108. Copyright 2020 by the author(s).

1 Introduction

Technological advances in data acquisition, management and warehousing (*e.g.* IoT, distributed platforms) enable massive data processing and are leading to a wide variety of new applications in the digitalized (service) industry. The need to design more and more automated systems fed by ever more informative streams of data manifests in many areas of human activity (*e.g* transportation, energy, health, commerce, finance/insurance). Monitoring the behavior/health of complex systems offers a broad spectrum of machine-learning implementation as classification or anomaly detection. With the increasing industrial digitalization, data are more and more often collected in quasi-real time and naturally take the form of temporal series or functions. The case of functional data is thus of crucial interest in practice, refer to *e.g.* Ramsay and Silverman (2002, 2005) for an excellent account of Functional Data Analysis (FDA in abbreviated form) and of its applications. A functional dataset is typically a set of $n \geq 1$ curves partially observed at different time points $T_1 = t_1 < \ldots < t_p = T_2$ which can be seen as n (partially observed) realizations of a stochastic process $X = (X_t)_{t\in[T_1,T_2]}$. Hence, the first step of FDA generally consists in reconstruct the functional objects from these observations, by means of interpolation, smoothing or projection techniques. Whereas, with the design of many successful algorithms such as (deep) neural networks, SVM's or boosting procedures, the practice of statistical learning has rapidly generalized these last few years, the automatic analysis of functional data to achieve complex (*e.g.* unsupervised) tasks such as anomaly detection is still a challenge, due to the huge variety of possible patterns that may carry the relevant information for discrimination purposes. It is far from straightforward to generalize directly methods originally introduced in the finite-dimensional case to the functional setup, unless preliminary *filtering* or dimensionality reduction techniques are used, see *e.g* Rossi and Villa (2006); Staerman et al. (2019). Such techniques essentially consist in projecting the observation, supposed to take their values in a certain Hilbert space,

onto a subspace of finite dimensionality, generally defined by truncating their expansion in a Hilbertian basis of reference or by means of a flexible dictionary of functions/'atoms'. Next, one can apply any state-of-the-art algorithm tailored to the finite dimension case, based on the parsimonious representations thus obtained, *cf* Ferraty and Vieu (2006); Ramsay and Silverman (2002). The basis functions can be either chosen among a pre-selected dictionary (*e.g.* Fourier, wavelets, cosine packets, *etc.*) presumably capable of capturing the information carried by the curves or built empirically using Functional Principal Component Analysis, retaining the most informative part of the (Kahrunen-Loève) decomposition only, see Ramsay and Silverman (2005). Of course, the major drawback of such FDA approaches lies in the fact that they are dramatically conditioned by the finite-dimensional representation method chosen, the subsequent analysis of the data may be fully jeopardized if the latter makes disappear some patterns relevant for the task considered.

Originally introduced by J. Tukey to extend the notion of median/quantile to multivariate random variables (see Tukey (1975a)), a *data depth* is a function defined on the feature space and valued in $[0, 1]$ used to determine the 'representativeness' of a point with respect to a statistical population and that should fulfill a variety of desirable properties (ideally just like the quantile function in the univariate situation). Given a training set, a data depth function provides a score that measures the centrality of any element w.r.t. a dataset and then defines, notably, an ordering of all the elements of the dataset. In particular, it finds a very natural application in (unsupervised) anomaly detection, see *e.g* Cuevas et al. (2007); Long and Huang (2016); Mosler and Mozharovskyi (2017) in supervised situations and Hubert et al. (2015); Nagy et al. (2017) in the unsupervised context: an observation is considered all the more 'abnormal', as its depth is small. This concept has been extended to the functional data framework by integrating univariate depth functions on the whole interval of time $[T_1, T_2]$, see *e.g.* Claeskens et al. (2014); Fraiman and Muniz (2001); Hubert et al. (2015). Alternatively, various depth functions fully tailored to the functional setup have been introduced in the statistical literature, refer to *e.g* Chakraborty and Chaudhuri (2014); Dutta et al. (2011); Lopez-Pintado and Romo (2009, 2011). However, most of them fail to fulfill certain desirable properties or face significant computational difficulties. It is the major purpose of this paper to introduce a novel robust functional statistical depth measure dedicated to the analysis of functional data. Based on the area of the convex hull of collections of sampled curves, it is easy to compute and to interpret both at the same time. Given a curve x lying in a certain functional space $\mathcal{F}$ (*e.g.* the Kolmogorov space $\mathcal{C}([0, 1])$, the space of real valued continuous functions on $[0, 1]$), the general idea is to quantify its contribution, on average, to the area of the convex hull (ACH in short) of random curves in $\mathcal{F}$ with the same probability law. Precisely, this function, referred to as the *ACH depth* throughout the article, is defined by the ratio of the ACH of the sample to that of the sample augmented by the curve x. We prove here that it fulfills various properties desirable for depth functions. In particular, given its form, it exhibits *sensitivity* (*i.e.* the depth score of new/test curves that are further and further away from the training set of curves decreases smoothly), which property, quite desirable intuitively, is actually not satisfied by most statistical (functional) depth documented in the literature. For instance, the statistical Tukey depth assigns a score of 0 to any element lying outside the convex hull of the training data, see Tukey (1975b); Dutta et al. (2011). In addition, the statistical depth we promote here is robust to outliers: adding outliers to the training set of curves has little or no effect on the returned score and ordering on a test set. For this reason, this functional depth is very well suited for unsupervised anomaly detection. In the functional setup, this task is extremely challenging. Indeed, the richness of functional spaces leads to a huge diversity in the nature of possibly observed differences between curves. As dicussed in Hubert et al. (2015), three main types of anomaly can be distinguished: *shift*, *magnitude* or *shape* anomalies. Anomalies can be either *isolated/transient* or *persistent* depending of their duration ; some of them being more difficult to detect (*shape* anomalies). Since the functional statistical depth measure we propose is related to a whole batch of curves and do not reflect the individual properties of a single curve, it enables the detection of a wide variety of anomaly shapes, as illustrated by the numerical experiments displayed in Section 4.

The paper is organized as follows. In Section 2, basic concepts pertaining to the statistical depth theory, both for the multivariate framework and for the functional case, are briefly recalled for clarity's sake. In Section 3 the functional statistical depth based on the area of the convex hull of a batch of curves is introduced at length and its theoretical properties are investigated, together with computational aspects. Section 4 presents numerical results in order to provide strong empirical evidence of the relevance of the novel depth function proposed, for the purpose of unsupervised functional anomaly detection especially. Eventually, concluding remarks are collected in Section 5.

2 Background and Preliminaries

For clarity, we start with recalling the concept of *statistical depth* in the multivariate and functional framework. We next list the desirable properties it should fulfill and also briefly review recent advances in this field. Here and throughout, the Dirac mass at any point a is denoted by δ_a, the convex hull of any subset A of $\mathbb{R}^d$ by $conv(A)$.

2.1 Data Depth in a Multivariate Space

By *data depth*, one usually means a nonparametric statistical function that determines the centrality of any element $x \in \mathbb{R}^d$ with respect to a statistical population. Given a dataset, a depth function provides a center-outward ordering of the data points. Since it permits to define a ranking of the (multivariate) observations and *local averages* derived from it, a *data depth* can be used for various tasks, including classification (Lange et al. (2014)), clustering (Jörnsten (2004)), anomaly detection (Serfling (2006)) or rank tests (Oja (1983)). In order to give a precise definition, some notations are needed. Let X be a random variable, defined on a certain probability space $(\Omega, \mathcal{A}, \mathbb{P})$, taking its values in $\mathcal{X} \subset \mathbb{R}^d$ with probability distribution P. Denote by $\mathcal{P}(\mathcal{X})$ the set of all probability distributions on $\mathcal{X}$. A data depth is a function

$$\begin{array}{rccc} D: & \mathbb{R}^d \times \mathcal{P}(\mathcal{X}) & \longrightarrow & [0,1] \\ & (x,\ P) & \longmapsto & D(x,\ P) \end{array}$$

measurable with respect to its first argument x. It is interpreted as follows: the closer the quantity $D(x, P)$ to 1, the deeper (*i.e.* the more 'central') the observation x with respect to the distribution P is considered. As mentioned above, it naturally defines a predorder on the set $\mathcal{X}$. In particular, medians of the multivariate distribution P corresponds to maximizers of the depth function D and quantile regions are defined as depth sublevel sets. A crucial example is the *half-space* depth D_T (also called *location* depth sometimes) introduced in the seminal contribution Tukey (1975b). It is defined as

$$D_T(x,\ P) = \inf\{P(H) :\ H \text{ closed half-space}, x \in H\},$$

for any $x \in \mathbb{R}^d$ and probability distribution on $\mathbb{R}^d$. As the distribution P is generally unknown, a statistical version can be built from independent copies $X_1, \ldots, X_n$ of the generic random vector X by means of the *plug-in* principle, *i.e.* by replacing P by an empirical counterpart $\widehat{P}_n$, typically the raw empirical distribution $(1/n)\sum_{i=1}^n \delta_{X_i}$ (or a smooth/penalized version of the latter), yielding the empirical depth

$$\widehat{D}_n(x) = D(x,\ \widehat{P}_n). \tag{1}$$

Empirical medians and quantile regions are then naturally defined as medians and quantile regions of the empirical depth (1). Of course, the relevance of a depth function regarding the measurement of *centrality* in a multivariate space is guaranteed in the sole case where certain desirable properties are satisfied. We refer to Zuo and Serfling (2000) for an account of the statistical theory of multivariate data depth and many examples.

2.2 Statistical Functional Depth

In this paper, we consider the situation where the r.v. X takes its values in a space of infinite dimension. Precisely, focus is on the case where the feature space is the vector space $\mathcal{C}([0,1])$ of real-valued continuous functions on $[0,1]$:

$$\begin{array}{rccccc} X & : & (\Omega,\ \mathcal{A},\ \mathbb{P}) & \longrightarrow & \mathcal{C}([0,1]) \\ & & \omega & \longmapsto & X(\omega) = (X_t(\omega))_{t\in[0,1]}. \end{array}$$

Recall that, when equipped with the sup norm $||.||_\infty$, $\mathcal{C}([0,1])$ is a separable Banach space. We denote by $\mathcal{P}(\mathcal{C}([0,1]))$ the set of all probability laws on $\mathcal{C}([0,1])$ and by P_t the 1-dimensional marginal of the law P of the stochastic process X at time point $t \in [0,1]$.

Depths in a functional framework have been first considered in Fraiman and Muniz (2001), where it is proposed to define functional depths as simple integrals over time of a univariate depth function D, namely $(x, P) \in \mathcal{C}([0,1]) \mapsto \int_0^1 D(x_t, P_t)dt$. Due to the averaging effect, local changes for the curve x only induce slight modifications of the depth value, which makes anomaly detection approaches based on such 'poorly sensitive' functional depths ill-suited in general. Recently, alternative functional depths have been introduced, see Lopez-Pintado and Romo (2009, 2011) for depths based on the geometry of the set of curves, Chakraborty and Chaudhuri (2014) for a notion of depth based on the L_2 distance or Dutta et al. (2011) for a functional version of the Tukey depth. As discussed in Nieto-Reyes and Battey (2016) and Gijbels and Nagy (2018), the axiomatic framework introduced in Zuo and Serfling (2000) for multivariate depth is no longer adapted to the richness of the topological structure of functional spaces. Indeed, the vast majority of the functional depths documented in the literature do not fulfill versions of the most natural and elementary properties required for a depth function in a multivariate setup, *cf* Gijbels and Nagy (2018). However, there is still no consensus about the set of desirable properties that a functional depth should satisfy, beyond the form of sensitivity mentioned above. Those that appear to be the most relevant in our opinion are listed below. By P_X is meant the law of a functional r.v. taking its values in $\mathcal{C}([0,1])$.

- (NON-DEGENERACY) For all non atomic distribution P in $\mathcal{P}(\mathcal{C}([0,1]))$, we have
$$\inf_{x\in\mathcal{C}([0,1])} D(x,P) < \sup_{x\in\mathcal{C}([0,1])} D(x,P).$$
- (AFFINE INVARIANCE) The depth D is said to be (scalar-) affine invariant if for any x in $\mathcal{C}([0,1])$ and all a, b in $\mathbb{R}$, we have
$$D(x,P_X) = D(ax+b, P_{aX+b}).$$
- (MAXIMALITY AT THE CENTER) For any point-symmetric and non atomic distribution P with $\theta \in \mathcal{C}([0,1])$ as center of symmetry, we have
$$D(\theta,P) = \sup_{x\in\mathcal{C}([0,1])} D(x,P).$$
- (VANISHING AT ∞) For any non atomic distribution P in $\mathcal{P}(\mathcal{C}([0,1]))$,
$$D(z,P) \underset{||z||_\infty\longrightarrow\infty}{\longrightarrow} \inf_{x\in\mathcal{C}([0,1])} D(x,P).$$
- (DECREASING W.R.T. THE DEEPEST POINT) For any P in $\mathcal{P}(\mathcal{C}([0,1]))$ such that $D(z,P) = \sup_{x\in\mathcal{C}([0,1])} D(x,P)$, $D(x,P) < D(y,P) < D(z,P)$ holds for any $x, y \in \mathcal{C}([0,1])$ such that $\min\{d(y,z), d(y,x)\} > 0$ and $\max\{d(y,z), d(y,x)\} < d(x,z)$.
- (CONTINUITY IN x) For any non atomic distribution $P \in \mathcal{P}(\mathcal{C}([0,1]))$, the function $x \mapsto D(x,P)$ is continuous w.r.t. the sup norm.
- (CONTINUITY IN P) For all x in $\mathcal{C}([0,1])$, the mapping $P \in \mathcal{P}(\mathcal{C}([0,1])) \mapsto D(x,\ P)$ is continuous w.r.t. the Lévy-Prohorov metric.

Before introducing the ACH depth and investigating its properties, a few remarks are in order. Though it obviously appears as mandatory to make the other properties meaningful, *non-degeneracy*, is actually not fulfilled by all the functional depths proposed, see *e.g* Lopez-Pintado and Romo (2009, 2011); Dutta et al. (2011). The '*Maximality at center*' and '*Decreasing w.r.t. the deepest point*' properties permit to preserve the original center-outward ordering goal as well as the goodness of fit for unimodal data of *data depth* in the functional framework. Many definition of the concept of "symmetry" in a functional space are detailed in the Supplementary material for the sake of place. The '*Continuity in x*' property extends a property fulfilled by cumulative distribution functions of multivariate continuous distributions. From a statistical perspective, the '*Continuity in P*' property is essential, insofar as P must be replaced in practice by an estimator, *cf* Eq. (1), built from finite-dimensional observations, *i.e.* a finite number of sampled curves.

3 The Area of the Convex Hull of (Sampled) Curves

It is the purpose of this section to present at length the statistical depth function we propose for path-valued random variables. As shall be seen below, its definition is based on very simple geometrical ideas and various desirable properties can be easily checked from it. Statistical and computational issues are also discussed at length. By $\mathcal{K}_2$ is meant the collection of all compact subsets of $\mathbb{R}^2$ and λ denotes Lebesgue measure on the plane $\mathbb{R}^2$. Consider an i.i.d. sample $X_1, \ldots, X_n$ drawn from P in $\mathcal{P}(\mathcal{C}([0,1]))$. The graph of any function x in $\mathcal{C}([0,1])$ is denoted by
$$graph(x) = \{(t,y) : y = x(t), t \in [0,1]\},$$
while we denote by $graph(\{x_1,\ldots,x_n\})$ the set
$$\bigcup_{i=1}^n graph(\{x_i\})$$
defined by a collection of $n \geq 1$ functions $\{x_1,\ldots,x_n\}$ in $\mathcal{C}([0,1])$. We now give a precise definition of the statistical depth measure we propose for random variables valued in $\mathcal{C}([0,1])$.

Definition 3.1 *Let $J \geq 1$ be a fixed integer. The ACH depth of degree J is the function $D_J : \mathcal{C}([0,1]) \times \mathcal{P}(\mathcal{C}([0,1])) \to [0,1]$ defined by: $\forall x \in \mathcal{C}([0,1])$,*
$$D_J(x,\ P) = \mathbb{E}\left[\frac{\lambda\left(conv\left(graph\left(\{X_1,\ldots,X_J\}\right)\right)\right)}{\lambda\left(conv\left(graph\left(\{X_1,\ldots,X_J\}\cup\{x\}\right)\right)\right)}\right],$$
where $X_1, \ldots, X_J$ are i.i.d. r.v.'s drawn from P. Its average version $\bar{D}_J$ is defined by: $\forall x \in \mathcal{C}([0,1])$,
$$\bar{D}_J(x,\ P) = \frac{1}{J}\sum_{j=1}^J D_j(x,\ P).$$

The choice of J leads to various views of distribution P, the average variant permitting to combine all of them (up to degree J). When $n \geq J$, an unbiased statistical estimation of $D_J(.,\ P)$ can be obtained by computing the symmetric U-statistic of degree J, see Lee (1990): $\forall x \in \mathcal{C}([0,1])$,
$$D_{J,n}(x) = \frac{1}{\binom{n}{J}} \sum_{1\leq i_1<\ldots<i_J\leq n} \frac{\lambda\left(conv\left(graph\left(\{X_{i_1},\ldots,X_{i_J}\}\right)\right)\right)}{\lambda\left(conv\left(graph\left(\{X_{i_1},\ldots,X_{i_J},x\}\right)\right)\right)}. \quad (2)$$
Considering the empirical average version given by
$$\forall x \in \mathcal{C}([0,1]),\quad \bar{D}_{J,n}(x) = \frac{1}{J}\sum_{j=1}^J D_{j,n}(x)$$

brings some 'stability'. However, the computational cost rapidly increasing with J, small values of J are preferred in practice. Moreover, as we illustrate in Section 4.1, J equal two already yields satisfactory results.

Approximation from sampled curves. In general, one does not observes the batch of continuous curves $\{X_1, \ldots, X_n\}$ on the whole time interval $[0,1]$ but at discrete time points only, the number $p \geq 1$ of time points and the time points $0 \leq t_1 < t_2 < \ldots < t_p \leq 1$ themselves possibly varying depending on the curve considered. In such a case, the estimators above are computed from continuous curves reconstructed from the sampled curves available by means of interpolation procedures or approximation schemes based on appropriate basis. In practice, linear interpolation is used for this purpose with theoretical guarantees (refer to Theorem 3.2 below) facilitating significantly the computation of the empirical ACH depth, see subsection 4.3.

3.1 Main Properties of the ACH Depth

In this subsection, we study theoretical properties of the population version of the functional depths introduced above and next establish the consistency of their statistical versions. The following result reveals that, among the properties listed in the previous subsection, five are fulfilled by the (average) ACH depth function.

Proposition 3.1 *For all $J \geq 1$, the depth function D_J (respectively, $\bar{D}_J$) fulfills the following properties: 'non-degeneracy', 'affine invariance', 'vanishing at infinity', 'continuity in x' and 'continuity in P'. In addition, the following properties are not satisfied: 'maximality at center' and 'decreasing w.r.t. the deepest point'.*

Refer to the Appendix section for the technical proof. In a functional space, not satisfying *maximality at center* is not an issue. For instance, though the constant trajectory $y(t) \equiv 0$ is a center of symmetry for the Brownian motion, it is clearly not representative of this distribution. In contrast, *scalar-affine invariance* is relevant, insofar as it allows z-normalization of the functional data and *continuity in P* is essential to derive the consistency of $D_{J,n}$ (respectively, of $\bar{D}_{J,n}$), as stated below.

Theorem 3.1 *Let $J \geq 1$ and $X_1, \ldots, X_n$ be $n \geq J$ independent copies of a generic r.v. X with distribution $P \in \mathcal{P}(\mathcal{C}([0,1]))$. As $n \to \infty$, we have, for any $x \in \mathcal{C}([0,1])$, with probability one,*

$$|D_{J,n}(x) - D_J(x,\ P)| \to 0$$

and

$$\left|\bar{D}_{J,n}(x) - \bar{D}_J(x,\ P)\right| \to 0.$$

3.2 On Statistical/Computational Issues

As mentioned above, only sampled curves are available in practice. Each random curve X_i being observed at fixed time points $0 = t_1^{(i)} < t_2^{(i)} < \ldots < t_{p_i}^{(i)} = 1$ (potentially different for each X_i) with $p_i \geq 1$, we denoted by $X_1', \ldots, X_n'$ the continuous curves reconstructed from the sampled curves $(X_i(t_1^{(i)}), \ldots, X_i(t_{p_i}^{(i)}))$, $1 \leq i \leq n$, by linear interpolation. From a practical perspective, one considers the estimator $D'_{J,n}(x)$ of $D_J(x,\ P)$ given by the approximation of $D_{J,n}(x)$ obtained when replacing the X_i's by the X_i''s in (2). The (computationally feasible) estimator $\bar{D}'_{J,n}(x)$ of $\bar{D}_J(x,\ P)$ is constructed in a similar manner. The result stated below shows that this approximation stage preserves almost-sure consistency.

Theorem 3.2 *Let $J \leq n$. Suppose that, as $n \to \infty$,*

$$\delta = \max_{1 \leq i \leq n} \max_{2 \leq k \leq p_i} \left\{ t_{k+1}^{(i)} - t_k^{(i)} \right\} \to 0.$$

As $n \to \infty$, we have, for any $x \in \mathcal{C}([0,1])$, with probability one,

$$\left|D'_{J,n}(x) - D_J(x,\ P)\right| \to 0$$

and

$$\left|\bar{D}'_{J,n}(x) - \bar{D}_J(x,\ P)\right| \to 0.$$

Refer to the Appendix section for the technical proof. Given the batch of continuous and piecewise linear curves $X_1', \ldots, X_n'$, although the computation cost of the area of their convex hull is of order $O(p \log p)$ with $p = \max_i p_i$, that of the U-statistic $D'_{J,n}(x)$ (and *a fortiori* that of $\bar{D}'_{J,n}(x)$) becomes very expensive as soon as $\binom{n}{J}$ is large. As pointed out in Lopez-Pintado and Romo (2009), even if the choice $J = 2$ for statistics of this type, may lead to a computationally tractable procedure, while offering a reasonable representation of the distribution, varying J permits to capture much more information in general. For this reason, we propose to compute an *incomplete* version of the U-statistic $D'_{J,n}(x)$ using a basic Monte-Carlo approximation scheme with $K \geq 1$ replications: rather than averaging over all $\binom{n}{J}$ subsets of $\{1, \ldots, n\}$ with cardinality J to compute $D'_{J,n}(x)$, one averages over $K \geq 1$ subsets drawn with replacement, forming an *incomplete U-statistic*, see Enqvist (1978). The same approximation procedure can be applied (in a randomized manner) to each of the U-statistics involved in the average $\bar{D}'_{J,n}(x)$, as described in the Supplementary Material.

4 Numerical Experiments

From a practical perspective, this section explores certain properties of the functional depth proposed using simulated data. It also describes its performance compared with the state-of-the-art methods on (real) benchmark datasets. As a first go, we focus on the impact of the choice of the tuning parameter K, which rules the trade-off between approximation accuracy and computational burden and parameter J. Precisely, it is investigated through the stability of the ranking induced by the corresponding depths. We next investigate the robustness of the ACH depth (ACHD in its abbreviated form), together with its ability to detect abnormal observations of various types. Finally, the ACH depth is benchmarked against alternative depths standing as natural competitors in the functional setup using real datasets. A simulation-based study of the variance of the ACH depth is postponed to the Supplementary Material.

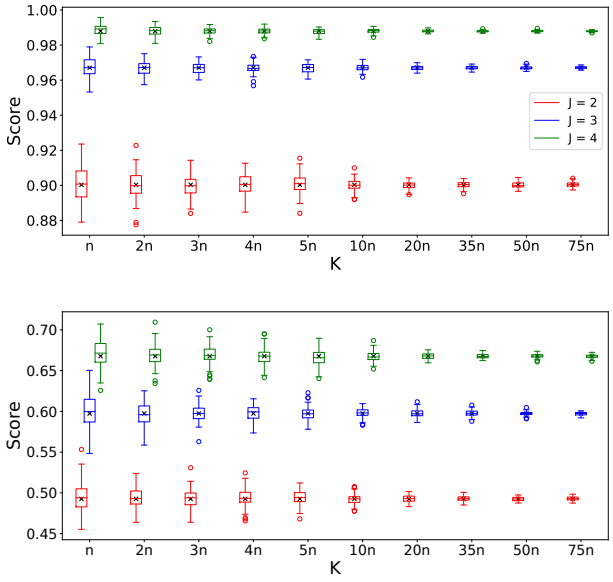


Figure 1: Boxplots of the approximations of $D_{J,n}(x_0)$ (top) and $D_{J,n}(x_3)$ (bottom) over different size of K. The black crosses correspond to the exact depth measure $D_{J,n}$ for each J respectively.

For the sake of simplicity, the two same simulated datasets, represented in Figure 3, are used throughout the section. The dataset (a) corresponds to sample path segments of the *geometric Brownian motion* with mean 2 and variance 0.5, a stochastic process widely used in statistical modeling. The dataset (b) consists of smooth curves given by $x(t) = a\cos(2\pi t) + b\sin(2\pi t)$, $t \in [0,1]$, where a and b are independently and uniformly distributed on $[0, 0.05]$, as proposed by Claeskens et al. (2014). Four curves $\{x_i : i \in \{0,1,2,3\}\}$ have been incorporated to each dataset: a deep curve and three atypical curves (*anomalies*), with expected depth-induced ranking $D_J(x_3) < D_J(x_2) \approx D_J(x_1) < D_J(x_0)$.

4.1 Choosing Tuning Parameters K and J

Parameter K reflects the trade-off between statistical performance and computational time. In order to investigate its impact on the stability of the method, we compute depths of the deepest and most atypical curves (x_0 and x_3) for dataset (b), taking $J = 2, 3, 4$. Figure 1 presents boxplots of the approximated ACHD (together with the exact values of ACHD) over 100 repetitions. Note that, as expected, depth values grow with J. The variance of the depth decreases taking sufficiently small values for $K = 5n$ and almost disappearing for $K \geq 20n$, while decreasing pattern remains the same for different values of K. For these reasons, we keep $K = 5n$ in what follows.

The choice of J is less obvious, and clearly when describing an observation in a functional space a substantial part of information is lost anyway. Nevertheless, one observes that computational burden increases exponentially with J and thus smaller values are preferable. Figure 2 shows the rank-rank plots of datasets (a) and (b) for small values of $J = 2, 3, 4$ and indicates, that depth-induced ranking does not change much with J. Thus, for saving computational time, we use value $J = 2$ in all subsequent experiments.

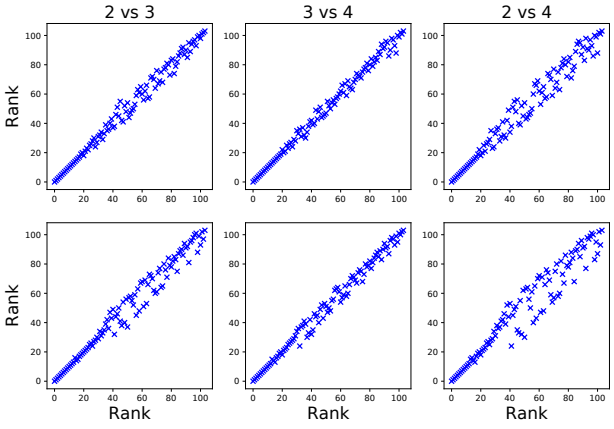


Figure 2: Rank-Rank plot for different values of J (2, 3 and 4). The first line represents the rank over the dataset (a) while the second line represents the dataset (b).

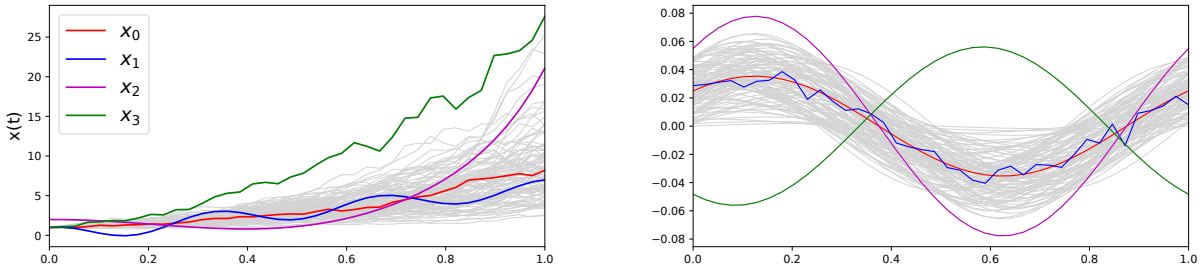


Figure 3: Datasets (a) (left) and (b) (right) containing 100 paths with four selected observations. The colors are the same for the four selected observations of both datasets (a) and (b).

$d_\tau(\sigma_0, \sigma_\alpha)(\times 10^{-2})$							
	$\boldsymbol{\alpha}$	0	5	10	15	25	30
ACHD	Location	0	**0.6**	**1.3**	**2.2**	**4.3**	**5.2**
	Isolated	0	**0.3**	**1.3**	**0.9**	**1.6**	**2.4**
	Shape	0	**0.9**	**2**	**2.6**	**4.2**	**4.7**
FSDO	Location	0	3.6	7.3	10	16	20
	Isolated	0	0.8	3.6	3.2	7.2	9.4
	Shape	0	1.6	2.9	4.2	6.6	7.4
FT	Location	0	5.1	9.5	13	20	23
	Isolated	0	0.7	2.7	2.7	5.9	7.2
	Shape	0	1.7	2.9	4.3	6.6	7.7
FIF	Location	0	7	8.2	7.3	7.3	8.9
	Isolated	0	9.3	12	11	10	12
	Shape	0	7.4	7.9	10	14	14

Table 1: Kendall's tau distances between the rank returned with normal data (σ_0) and contamined data (σ_α, over different portion of contamination α with location, isolated and shape anomalies) for ACHD and three state-of-the-art methods. Bold numbers indicate best stability of the rank over the contaminated datasets.

4.2 Robustness

Under robustness of a statistical estimator on understands its ability not to be "disturbed" by atypical observations. We explore robustness of ACHD in the following simulation study: between the original dataset and the same dataset contaminated with anomalies, we measure (averaged over 10 random repetitions) Kendall's τ distance of two depth-induced rankings σ and σ', respectively, of the original data:

$$d_\tau(\sigma, \sigma') = \frac{n(n-1)}{2} \sum_{i<j} \mathbb{I}_{\{(\sigma(i)-\sigma(j))(\sigma'(i)-\sigma'(j))<0\}} \cdot$$

In their overview work, Hubert et al. (2015) introduce taxonomy for atypical observations, focusing on *Location*, *Isolated*, and *Shape* anomalies. Here, we add *Location* anomalies to dataset (a) and *Isolated* and *Shape* anomalies to dataset (b); other types of anomalies for both datasets can be found in the Supplementary Material. The abnormal functions are constructed as follows. *Location* anomalies for dataset (a) are $\tilde{x}(t) = x(t) + ax(t)$ with a drawn uniformly on $[0, 1]$. *Isolated* anomalies for dataset (b) are constructed by adding a peak at t_0 (drawn uniformly on $[0, 1]$) of amplitude b (drawn uniformly on $[0.03, 0.06]$) such that $\tilde{y}(t_0) = y(t_0) + b$ and $\tilde{y}(t) = y(t)$ for any $t \neq t_0$. *Shape* anomalies for dataset (b) are $\tilde{z}(t) = z(t) + 0.01 \times \cos(2\pi t f) + 0.01 \times \sin(2\pi t f)$ with f drawn uniformly from $\{1, 2, ..., 10\}$. By varying the percentage of abnormal observations α, we compare ACHD to several of the most know in the literature depth approaches: the *functional Stahel-Donoho depth* (FSDO) (Hubert et al., 2015) and the *functional Tukey depth* (FT) (Claeskens et al., 2014), and also to the *functional isolation forest* (FIF) algorithm (Staerman et al., 2019) which proves satisfactory anomaly detection; see Table 1. One can observe that ACHD consistently preserves depth-induced ranking despite inserted abnormal observation, even if their fraction α reaches 30%. FSDO behaves competitively giving

slightly better results than ACHD for shape anomalies.

4.3 Applications to Anomaly Detection

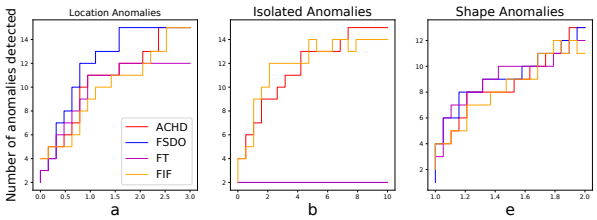


Figure 4: Number of anomalies detected over a grid of parameters for three types of anomalies (location, isolated, and shape) for ACHD and three further state-of-the-art methods.

Further, we explore the ability of ACHD to detect atypical observations. For this, we conduct an experiment in settings similar to those in Section 4.2, while changing degree of abnormality gradually for 15 (out of 100 curves) in dataset (a). Thus, we alter a in $[0, 3]$ for *location anomalies*, b in $[0, 10]$ for *isolated anomalies*, and e in $[1, 2]$ for *shape anomalies* to amplify the "spikes" of oscillations such that $\tilde{z}(t) = ez(t)$. (For an illustration of abnormal curves the reader is referred to the Supplementary Material.) Figure 4 illustrates number of anomalies detected by ACHD, FSDO, FT, and FIF for different parameters of abnormality. While it is difficult to find the general winner, ACHD behaves favorably in all the considered cases and clearly outperforms the two other depths when the data is contaminated with *isolated anomalies*.

We conclude this section with a *real-world data benchmark* based on three datasets: Octane (Esbensen, 2001), Wine (Larsen et al., 2006), and EOG (Chen et al., 2015). The Wine dataset consists of 397 measurements of proton nuclear magnetic resonance (NMR) spectra of 40 different wine samples, the Octane dataset are 39 near infrared (NIR) spectra of gasoline samples with 226 measurements, while the EOG dataset represents the electrical potential between electrodes placed at points close to the eyes with 1250 measurements. (Graphs of the three datasets can be found in the Supplementary Material.) As pointed out by Hubert et al. (2015), it is difficult to detect anomalies in the first two datasets, while they are easily seen during the human eye inspection. For the EOG dataset, we assign smaller of the two classes to be abnormal. To the existing state-of-the-art methods, we add here *Isolation Forest* (IF) (Liu et al., 2008) and the *One-Class SVM* (OC) (Schölkopf et al., 2001)—multivariate methods applied after a proper dimension reduction (to the dimension 10) using *Functional Principal Component Analysis* (FPCA).(Ramsay and Silverman, 2002). Portions of detected anomalies (by all the considered methods), indicated in Table 2, hint on very competitive performance of ACHD in the addressed benchmark.

	ACHD	FSDO	FT	FIF	IF	OC
Octane	**1**	0.5	0.33	**1**	0.5	0.5
Wine	**1**	0	0	**1**	0	**1**
EOG	**0.73**	0.55	0.48	0.43	0.63	0.6

Table 2: Portion of detected anomalies of benchmark methods for the Octane, Wine, and EOG datasets.

5 Conclusion

In this paper, we have introduced a novel functional depth function on the space $\mathcal{C}([0, 1])$ of real valued continuous curves on $[0, 1]$ that presents various advantages. Regarding interpretability first, the depth computed at a query curve x in $\mathcal{C}([0, 1])$ takes the form of an expected ratio, quantifying the relative increase of the area of the convex hull of i.i.d. random curves when adding x to the batch. We have shown that this depth satisfies several desirable properties and have explained how to solve approximation issues, concerning the sampled character of observations in practice and scalability namely. Numerical experiments on both synthetic and real data have highlighted a number of crucial benefits: reduced variance of its statistical versions, robustness with respect to the choice of tuning parameters and to the presence of outliers in the training sample, capacity of detecting (possibly slight) anomalies of various types, surpassing competitors such as depths of integral type, for isolated anomalies in particular. The open-source implementation of the method, along with all reproducing scripts, can be accessed at https://github.com/GuillaumeStaermanML/ACHD.

Acknowlegments

The authors thank Stanislav Nagy for his helpful remarks. This work has been funded by BPI France in the context of the PSPC Project Expresso (2017-2021).

References

Chakraborty and Chaudhuri (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The annals of statistics*.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. (2015). The ucr time series classification archive.

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of American Statistical Association*, 109(505):411–423.

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

Dutta, Ghosh, and Chaudhuri (2011). Some intriguing properties of tukey's half-space depth. *Bernouilli*.

Enqvist, E. (1978). *On sampling from sets of random variables with application to incomplete U-statistics*. PhD thesis, Lund University.

Esbensen, K. (2001). Multivariate data analysis-in practice. *Camo Software*.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis*. Springer-Verlag, New York.

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

Gijbels, I. and Nagy, S. (2018). On a general definition of depth for functional data. *Statistical Science*.

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications*, 24(2):177–202.

Jörnsten, R. (2004). Clustering and classification based on the l1 data depth. *Journal of Multivariate Analysis*, 90(1):67 – 89.

Lange, T., Mosler, K., and Mozharovskyi, P. (2014). Fast nonparametric classification based on data depth. *Statistical Papers*, 55(1):49–69.

Larsen, F., Berg, F., and Engelsen, S. (2006). An exploratory chemometric study of 1 h nmr spectra of table wine. *Journal of Chemometrics*, 20:198 – 208.

Lee, A. J. (1990). *U-statistics: Theory and practice*. Marcel Dekker, Inc., New York.

Liu, F. T., Ting, K. M., and Zhou, Z. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422.

Long and Huang (2016). A study of functional depths. *preprint*.

Lopez-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*.

Lopez-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Computational Statistics and Data Analysis*.

Mosler, K. and Mozharovskyi, P. (2017). Fast dd-classification of functional data. *Statistical Papers*, 58(4):1055–1089.

Nagy, S., Gijbels, I., and Hlubinka, D. (2017). Depth-based recognition of shape outlying functions. *Journal of Computational and Graphical Statistics*, 26(4):883–893.

Nieto-Reyes, A. and Battey, H. (2016). A topologically valid definition of depth for functional data. *Statistical Science*.

Oja (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag, New York.

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

Rossi, F. and Villa, N. (2006). Support vector machine for functional data classification. *Neurocomputing*, 69(7):730 – 742.

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., and Williamson, R. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471.

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

Staerman, G., Mozharovskyi, P., Clémençon, S., and d'Alché Buc, F. (2019). Functional isolation forest. In *Proceedings of The 11th Asian Conference on Machine Learning*.

Tukey (1975a). Mathematics and the picturing of data. *In Proceedings of the International Congress of Mathematicians*.

Tukey, J. W. (1975b). Mathematics and the picturing of data. In James, R., editor, *Proceedings of the International Congress of Mathematicians*, volume 2, pages 523–531. Canadian Mathematical Congress.

Zuo, B. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.