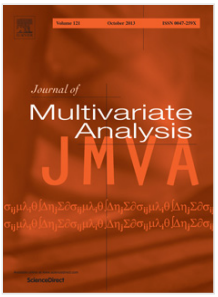

An angle-based multivariate functional pseudo-depth for shape outlier detection

S. Kuhnt, A. Rehage

PII: S0047-259X(15)00267-5
DOI: http://dx.doi.org/10.1016/j.jmva.2015.10.016
Reference: YJMVA 4039

To appear in: *Journal of Multivariate Analysis*

Received date: 31 January 2015

Please cite this article as: S. Kuhnt, A. Rehage, An angle-based multivariate functional pseudo-depth for shape outlier detection, *Journal of Multivariate Analysis* (2015), http://dx.doi.org/10.1016/j.jmva.2015.10.016

An angle-based multivariate functional pseudo-depth for shape outlier detection

S. Kuhnt[a], A. Rehage[b,*]

[a]Department of Computer Sciences, Dortmund University of Applied Sciences and Arts, 44227 Dortmund, Germany
[b]Faculty of Statistics, TU Dortmund University, 44221 Dortmund, Germany

Abstract

A measure especially designed for detecting shape outliers in functional data is presented. It is based on the tangential angles of the intersections of the centred data and can be interpreted like a data depth. Due to its theoretical properties we call it functional tangential angle (FUNTA) pseudo-depth. Furthermore we introduce a robustification (rFUNTA). The existence of intersection angles is ensured through the centring. Assuming that shape outliers in functional data follow a different pattern, the distribution of intersection angles differs. Furthermore we formulate a population version of FUNTA in the context of Gaussian processes. We determine sample breakdown points of FUNTA and compare its performance with respect to outlier detection in simulation studies and a real data example.

Keywords: Bootstrap, Data depth, Functional data, Robust estimate, Shape outlier detection
2010 MSC: 60G15, 62H30, 62P99, 68-04

1. Introduction

In every statistical analysis one should check the data for unusual, surprising observations. In one or two dimensions this might be done graphically. However, in higher dimensions or complex data structures, outlier detection methods are needed [4].

The analysis of outliers in functional data [33, 14, 20, 6] is an interesting and recent field. Functional data occur when measurements at each observational unit are taken densely, mostly over time. The inherent smoothness of the measurements is used to approximate the data vector (or matrix) by a function. As in a regression setup (see e.g. [24]), in functional data it can be distinguished between different types of outliers. For example, magnitude and shape outliers are mentioned in [27]. A finer classification is revisited in [21]: Shift outliers are curves which are shifted away from the bulk of the data, whereas amplitude outliers are curves with a similar shape but differing scale. Furthermore the authors differentiate between persistent and isolated outliers. However, an isolated outlier with e.g. a larger scale than the remaining data might also be seen as a global (persistent) shape outlier. We focus on shape outliers, since shift and magnitude outliers can often be detected with well-known univariate approaches. A popular approach to apply a centre-outward ordering to multivariate as well as functional data is the concept of data depth. Several outlier identification approaches based on functional depth measures exist [9, 12, 27], but they are not specifically designed to detect shape outliers. See for example Figure 1, where the dashed functions have the same modified band depth on each of the panels. In the left panel the solid functions have a larger depth which is sensible for magnitude outlier detection, but not for shape outlier detection. In the right panel one of the dashed functions follows an obviously different pattern than the rest, but that is not reflected in the modified band depth. Recently, a graphical approach [3] has been proposed to overcome this issue. We try to remedy this shortcoming more directly by introducing a measure depending on the angles in the intersections of one observation with the others, which we call FUNTA pseudo-depth or to put it briefly: FUNTA. Similar to the location-slope graph depth [29], FUNTA depends on functions and their derivatives, but restricts the analysis to the intersections.

*Corresponding author
Email addresses: sonja.kuhnt@fh-dortmund.de (S. Kuhnt), rehage@statistik.tu-dortmund.de (A. Rehage)

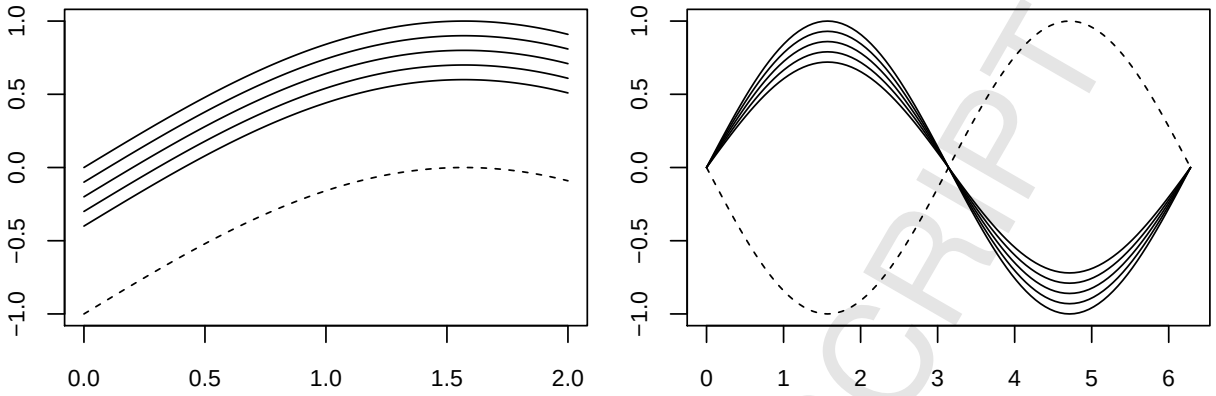

Figure 1: Toy data examples where modified band depth [27] yields implausible results.

As in the non-functional case, one encounters not only univariate, but also multivariate data. Literature regarding the so-called multi-functional data analysis emerged recently, expanding methods like linear models [33], partial linear modelling [2], additive prediction and boosting [15] as well as linked applications like variable selection [8] to the multi-functional context. Besides, the concept of data depth has been extended to multi-functional situations: One way to deal with multivariate functional data is to compute one of the numerous multivariate data depths for each time point. This has been conducted for univariate and multivariate functional data in [21] using the skew-adjusted projection depth and a heat map. While in the univariate situation outliers could be found easily, the multivariate situation has turned out to be more ambiguous. Hence, the authors use a further graphical representation to clarify multivariate functional outliers in a dataset, called the centrality-stability plot. On the other hand, multivariate functional depths have been constructed to overcome this issue, for example the multivariate functional halfspace depth [7, 22], the simplicial band depth for multivariate functional data [28] and a modification using weighting variance-covariance operators [5]. As a competitor to those depths, we introduce a multivariate version of FUNTA along with the average distance between curves in this paper.

An alternative framework for outlier detection is described in [14]: In functional nonparametric unsupervised classification, the outlying and non-outlying data can be considered as two (or more) classes. The advantage of our approach is that the number of classes is irrelevant, hence the outliers do not need to be homogeneous in any way.

It is well-known that "for functional data the information contained in the *shape* of the curves matters a great deal" [20, p. 5]. Thus it is surprising that there are barely any depth measures focusing on the curves' shape. Certainly, depth measures were intended to determine a multivariate median and outlier detection is in various aspects different to median detection. Particularly in functional data situations we emphasize that a shape outlier is not necessarily the "least median" observation. To improve the ability of a measure to detect shape outliers, any information concerning the location should be removed from the data beforehand. In chemometrics, this procedure is known as the baseline correction (see e.g. [21]). Therefore we will mainly refer to "centred functions" in the following section.

The paper is organised as follows. In Section 2 we introduce two different notions of the functional tangential angle pseudo-depth, both univariate and multivariate. A population version is stated in the case of Gaussian processes. We also discuss theoretical properties of FUNTA, particularly breakdown points. We compare the ability to detect outliers in different artificial data situations in Section 3. Besides the two versions of FUNTA we choose h-modal depth, modified band depth, the integrated depth and the multivariate functional halfspace depth as competitors. A real data example from linguistics is analysed in Section 4. Conclusions and projections for further research are given in Section 5.

2. Methodology

In this section we present the FUNTA measurement based on functional tangential angles and then introduce robustified, multivariate and population versions. Computational issues are discussed and sample breakdown properties derived.

2.1. Sample versions of FUNTA pseudo-depth

Before we define FUNTA in a slightly different notation than in [35], we state that by a sample of "centred functions" we mean that $\int x_i(t)dt = \int x_j(t)dt \; \forall \; i, j$.

Definition 1. Functional tangential angle pseudo-depth: FUNTA

Let $\tilde{x}(t), x_1(t), \ldots, x_n(t), t \in \mathcal{T} = [a, b]$, be a sample of real centred differentiable functions, where $\tilde{x}(t)$ has a finite number of intersections with $x_i(t), i = 1, \ldots, n$, denoted by $s_{i,k}, k = 1, \ldots, m_i$. The total number of intersections of $\tilde{x}(t)$ is given by $\sum_{i=1}^{n} m_i = m$. Let γ_k denote the geometric angle (radiant measure) of the tangent to $\tilde{x}(s_{i,k})$ with the tangent to $x_i(s_{i,k})$. Then we call

$$\text{FUNTA}_n(\tilde{x}) = 1 - (m\pi)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{m_i} \gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})) \tag{1}$$

functional tangential angle (FUNTA) pseudo-depth *of $\tilde{x}(t)$ w.r.t. the sample.*

The concept of the tangential angle in Eq. (1) is visualised in Fig. 2.

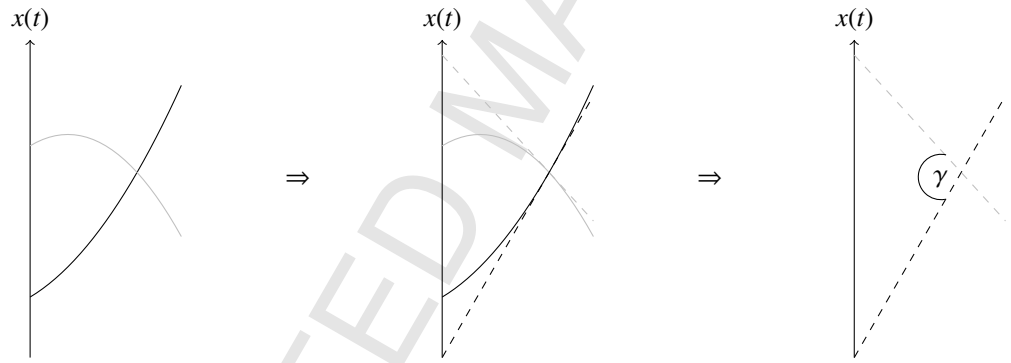

Figure 2: In left plot, two solid functions (x-axis omitted to improve readability) with one intersection are depicted. In the middle we added the tangents (dashed) of the functions in the intersection. On the right the intersection angle $\gamma \in [0, \pi)$ is shown. We use the radiant measure because it is dimensionless.

It is well-known that the intersection angle of two curves (functions) $\tilde{x}$ and x_i equals the intersection angle of the tangents (derivatives) in the intersections. In the situation of Def. 1, let $\tilde{x}'$ be the derivative of $\tilde{x}$. Then, the intersection angle is given by

$$\gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})) = \arccos\left(\frac{1 + \tilde{x}'(s_{i,k}) x_i'(s_{i,k})}{\sqrt{(1 + \tilde{x}'(s_{i,k})^2)(1 + x_i'(s_{i,k})^2)}} \right).$$

Remark 1. *The* FUNTA *pseudo-depth takes values in* $(0, 1]$.

Remark 1 can be seen easily by inserting the extremal values for intersection angles, i.e. $\gamma \in [0, \pi)$. The larger the average intersection angle is, the smaller becomes FUNTA. See Fig. 3 for an example: Three functions $x_1(t), x_2(t), x_3(t)$

are depicted, whereupon $x_3(t)$ follows a different pattern than the others, which results in a considerably smaller FUNTA. Now consider $rx_i(t), r = 2, i = 1, 2, 3$, which increases the disparity of $x_3(t)$ with respect to $x_1(t), x_2(t)$. Hence, $\text{FUNTA}_n(2x_3)$ is now even smaller than $\text{FUNTA}_n(x_3)$ in the previous situation. In the limit it holds that $\lim_{r\to\infty} \text{FUNTA}_n(rx_3) = 0$. On the other hand, an outlier-free sample (without x_3) yields $\lim_{r\to\infty} \text{FUNTA}_n(rx_1) = \lim_{r\to\infty} \text{FUNTA}_n(rx_2) = 1$.

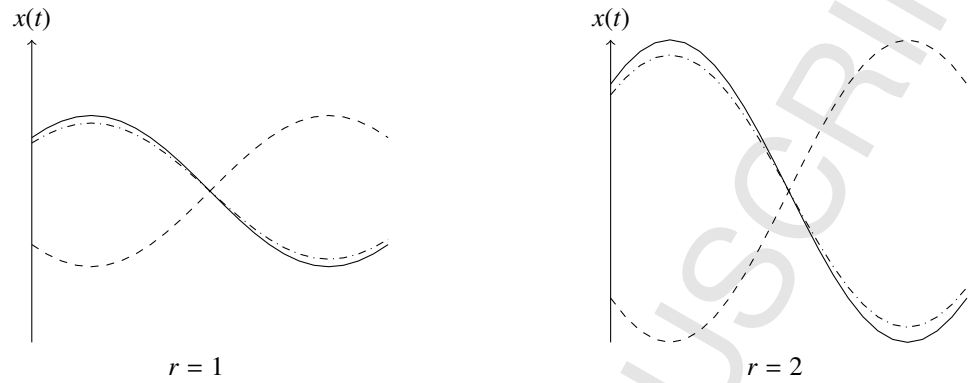

Figure 3: On the left panel the functions $x_1(t) = \sin(t)$ (solid), $x_2(t) = 0.9\sin(t)$ (dashdotted) and $x_3(t) = -\sin(t)$ (dashed) are depicted. Given the displayed domain $\mathcal{T} = [\pi/4, 7\pi/4]$, $\text{FUNTA}_n(x_1) = 0.621$, $\text{FUNTA}_n(x_2) = 0.627$, $\text{FUNTA}_n(x_3) = 0.506$. On the right panel the functions are multiplied by $r = 2$. This yields $\text{FUNTA}_n(2x_1) = 0.641$, $\text{FUNTA}_n(2x_2) = 0.648$, $\text{FUNTA}_n(2x_3) = 0.302$.

Nevertheless it is possible to gain a FUNTA value of one without considering the limit behaviour, e.g. for a sample $x_4(t) = t^3, x_5(t) = 2t^3, \mathcal{T} = [-1, 1]$, where $\text{FUNTA}_n(x_4) = \text{FUNTA}_n(x_5) = 1$.

Remark 2. *Definition 1 assumes a finite number of intersections. However, if complete areas of intersection (in the sense that $\tilde{x}(s) = x_i(s)\ \forall\ s \in [s_1, s_2], s_1 \neq s_2$) are present, we handle them as two intersections in $\{s_1, s_2\}$. Using this convention it is again possible to calculate* FUNTA *values.*

One can show that the existence of at least one intersection of two centred functions is ensured:

Proposition 1. *Let $x_1(t), x_2(t), t \in \mathcal{T} = [a, b]$ with $\int_a^b x_1(t)dt = \int_a^b x_2(t)dt$. Furthermore $x_i'(t) \in \mathbb{R}\ \forall\ t \in \mathcal{T}, i = 1, 2$. Then at least one intersection $\tilde{t} \in \mathcal{T}$ exists, such that $x_1(\tilde{t}) = x_2(\tilde{t})$.*

P . Let $f_{1,2}(t) := x_1(t) - x_2(t)$. It is easy to see that $f_{1,2}$ is a continuously differentiable function. Take an arbitrary point $d \in \mathcal{T}$. We then have three possible cases.

1. If $f_{1,2}(d) = 0$, d is an intersection.
2. If $f_{1,2}(d) < 0$, it holds that $x_1(d) < x_2(d)$. Because of the intermediate value theorem, it suffices to show the existence of $\tilde{d} \in [a, b]$ with $x_1(\tilde{d}) > x_2(\tilde{d})$. Suppose that no such $\tilde{d}$ exists, hence $x_1(\tilde{d}) \leq x_2(\tilde{d})$ for all $\tilde{d} \in [a, b]$. If $\tilde{d}$ exists with $x_1(\tilde{d}) = x_2(\tilde{d})$ this is an intersection. If no such $\tilde{d}$ exists, we have $x_1(\tilde{d}) < x_2(\tilde{d})\ \forall\ \tilde{d} \in [a, b]$, hence $\int_a^b x_1(t)dt < \int_a^b x_2(t)dt$, which is a contradiction to the assumption.
3. Let $d \in \mathcal{T}$ with $f_{1,2}(d) > 0$. Define $\tilde{f}_{1,2}(d) := -f_{1,2}(d)$ and conduct the steps of the second case to show the existence of an intersection.

Remark 3. *If each function of a sample $x_1(t), \ldots, x_n(t)$ is centred with its own mean, then it holds that $\int_a^b x_i(t)dt = 0\ \forall\ i = 1, \ldots, n$.*

In the subsequent propositions we see that FUNTA does not fulfill the properties of a statistical depth function [39] or functional depth [30, 31]. We like to remark that these properties are of course sensible for location-related ranking, but not all of them can be translated in a meaningful way to shape-specific issues.

Proposition 2. FUNTA *is invariant to a shift in centring: Denote* $\text{FUNTA}_n(x_1, \ldots, x_n)$ *as the vector containing* FUNTA *pseudo-depth values of the centred functions $x_1, \ldots, x_n$. Then* $\text{FUNTA}_n(x_1+l, \ldots, x_n+l) = \text{FUNTA}_n(x_1, \ldots, x_n)$ *holds for $l \in \mathbb{R}$.*

P . Since $x_i'(t) + l_i'(t) = x_i'(t)$, this is obvious.

Proposition 2 implies that Def. 1 is also valid for other common mean values. However, FUNTA is not invariant w.r.t. addition of some non-constant function $l(t)$ to every function in the sample. Consider the following example: $x_1(t) = t^2 - 1, x_2(t) = 1 - t^2$ for $\mathcal{T} = [-\sqrt{3}, \sqrt{3}]$. Then x_1, x_2 are centred on zero with two intersections $s_1 = -1, s_2 = 1$. Then $\gamma_1(x_1(-1), x_2(-1)) = \arccos\left(\frac{1-2\times2}{\sqrt{5}\times\sqrt{5}}\right) = \arccos(-3/5) = 2.214$. On the other hand, for $l(t) = t^3$, we have $\gamma_1(x_1(-1) + l(-1), x_2(-1) + l(-1)) = \arccos\left(\frac{1+1\times5}{\sqrt{2}\times\sqrt{26}}\right) = \arccos(6/\sqrt{52}) = 0.588$.

Proposition 3. FUNTA *is not scale equivariant:* $\exists\, s \in \mathbb{R}, x_1(t), x_2(t) : \text{FUNTA}_n(s\times x_1(t), s\times x_2(t)) \neq s\,\text{FUNTA}_n(x_1(t), x_2(t))$.

P . This follows from the fact that arccos is not a linear function.

Proposition 4. FUNTA *is not scale invariant:* $\exists\, s \in \mathbb{R}, x_1(t), x_2(t) : \text{FUNTA}_n(s\times x_1(t), s\times x_2(t)) \neq \text{FUNTA}_n(x_1(t), x_2(t))$.

P . Let $x_1'(1) = 1, x_2'(1) = 2$ have an intersection in $t = 1$. The outcome of this is the intersection angle $\gamma(x_1(1), x_2(1)) = 0.322$, but for $s = 3$ we get $\gamma(3x_1(1), 3x_2(1)) = 0.156$.

Hence FUNTA does not fulfill all requirements of a data depth as stated in [39]. However, its values can be interpreted just like those of a functional depth: A small depth value indicates shape outlyingness. We therefore propose to call such a measure pseudo-depth.

The FUNTA pseudo-depth as given in Def. 1 will be very sensitive to shape outliers having a higher number of intersections (and larger angles) than the bulk of the data. Not only the depth value of the outlier will be suspicious, but also many depth values of the regular data will be shifted from their value in a sample without the shape outlier. In order to prevent such a behaviour, we replace the mean in Eq. (1) by a more robust measure, namely the median of the maximum of the (function-)pairwise angles. By that FUNTA is robustified, since a small number of very extreme observations has barely no influence on the depth of other observations.

Definition 2. Robustified FUNTA pseudo-depth

Using the notation of Def. 1, we define the robustified FUNTA *pseudo-depth as*

$$\text{rFUNTA}_n(\tilde{x}) = 1 - \pi^{-1} \underset{i=1,\ldots,n}{\text{median}}(\max_k(\gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})))).$$

Another approach to robustify FUNTA could be the replacement of the mean angle by the median angle in Eq. (1). Unfortunately, such a measure would lose most of its sensitivity to detect shape outliers. Typically, a vector of intersection angles of a functional outlier itself contains a number of *non*-outlying values. Since this percentage is not necessarily smaller than 0.5, a median of this vector could fail to detect the difference.

Recently the development and analysis of data depths for *multivariate* functional data moved into focus [7, 28, 21]. In a multivariate functional data situation we observe a p-dimensional data vector at each point $t_i, i = 1, \ldots, T$. In contrast to that, $p = 1$ holds for usual univariate functional data situations. Particularly interesting multivariate outliers are those which are not outlying with respect to the marginal functional depths, but outlying with respect to the correlation structure of the dimensions. A nice example for this case is given in [28] by an analysis of the growth and weight of boys. One boy grows quite slowly and gains weight quickly. None of the two curves are suspicious on their own, but analysed by a suitable multivariate functional depth the outlyingness of this observation becomes visible. On the other hand, the focus on the correlation structure should not be too strong. For example, having correlated, yet considerable shape outliers in every dimension in a set of correlated multivariate functional data, the outlyingness should still be detected by the chosen depth measure. To solve this issue, the detection of marginal shape outliers could be conducted in a previous step. For the pre-step, the extensions of FUNTA and rFUNTA to multivariate functional data are straightforward.

Definition 3. Multivariate FUNTA pseudo-depth

For $i = 1, \ldots, n$, let $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,p})^T$ be the vector of functions belonging to the i^{th} observation and $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_p)^T$ an arbitrary vector of functions from the sample space. Furthermore we denote $s_{i,j,k}$ as the k^{th} intersection of $\tilde{x}_j$ with $x_{i,j}, k = 1, \ldots, m_{i,j}, \sum_{i=1}^{n} \sum_{j=1}^{p} m_{i,j} = m$. Then we define a multivariate FUNTA *pseudo-depth by*

$$\text{FUNTA}_n(\tilde{\mathbf{x}}) = 1 - (m\pi)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{m_{i,j}} \gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k})).$$

Due to the arithmetic mean operator, outlyingness in a small number of dimensions might get lost. To overcome this issue, we define the multivariate version of rFUNTA as follows.

Definition 4. Multivariate robustified FUNTA pseudo-depth
Using the notation of Def. 3, we define a multivariate robustified FUNTA *pseudo-depth by*

$$\text{rFUNTA}_n(\tilde{\mathbf{x}}) = 1 - \pi^{-1} \underset{i=1,\ldots,n}{\text{median}}(\max_j(\max_k(\gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k}))))).$$

Taking the maximum (w.r.t. the dimension) of the maximum (w.r.t. the intersecting curve) angles achieves the sensitivity to identify outliers.

Remark 4. *It is obvious that Propositions 2, 3 and 4 also hold for* rFUNTA. *Moreover, Prop. 2 holds for Definitions 2, 3, and 4.*

The following Prop. 5 illustrates why a further step has to be added to the multivariate outlier detection procedure.

Proposition 5. *The multivariate* FUNTA *is a weighted average of the marginal* FUNTA *pseudo-depths.*

P .

$$\begin{aligned}
&\text{FUNTA}_n(\tilde{x}_j) = 1 - (m_j\pi)^{-1} \sum_{i=1}^{n} \sum_{k=1}^{m_{i,j}} \gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k})) \\
\Leftrightarrow\ & m_j\pi(1 - \text{FUNTA}_n(\tilde{x}_j)) \overset{(*)}{=} \sum_{i=1}^{n} \sum_{k=1}^{m_{i,j}} \gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k})). \\
\Rightarrow\ & \text{FUNTA}_n(\tilde{\mathbf{x}}) = 1 - (m\pi)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{k=1}^{m_{i,j}} \gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k})) \\
&\qquad = 1 - (m\pi)^{-1} \sum_{j=1}^{p} \sum_{i=1}^{n} \sum_{k=1}^{m_{i,j}} \gamma_k(\tilde{x}_j(s_{i,j,k}), x_{i,j}(s_{i,j,k})) \\
&\qquad \overset{(*)}{=} m^{-1} \sum_{j=1}^{p} m_j \, \text{FUNTA}_n(\tilde{x}_j).
\end{aligned}$$

Following Proposition 5, multivariate FUNTA does not take a possible dependence structure between the functions into account. Since we already computed FUNTA, the intersection angles x_i has in dimension j_1 are known. We note that the number of intersection angles will—in practice—be different across the dimensions *and* across the observations. Therefore we cannot use the standard correlation idea. Consider these angles $\gamma_{1,j_1}, \ldots, \gamma_{m_{j_1},j_1}$ now in order of appearance. In dimension j_2, x_i has angles $\gamma_{1,j_2}, \ldots, \gamma_{m_{j_2},j_2}$. For simplicity, assume that the angles are equidistant on the unit scale, $\Im_1 = \{0, \frac{1}{m_{j_1}-1}, \frac{2}{m_{j_1}-1}, \ldots, 1\}$ and $\Im_2 = \{0, \frac{1}{m_{j_2}-1}, \frac{2}{m_{j_2}-1}, \ldots, 1\}$ respectively. In practice, this reduces the storage space of the algorithm and can be justified since the number of intersections is typically large enough that taking the actual intersection points instead of $\Im_i$ would barely make any difference. We connect the points $\left(\begin{pmatrix} 0 \\ \gamma_{1,j_1} \end{pmatrix}, \begin{pmatrix} 1/(m_{j_1}-1) \\ \gamma_{2,j_1} \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ \gamma_{m_{j_1},j_1} \end{pmatrix}\right)$ to a polygonal chain. For dimension j_2, we obtain a polygonal chain analogously. A

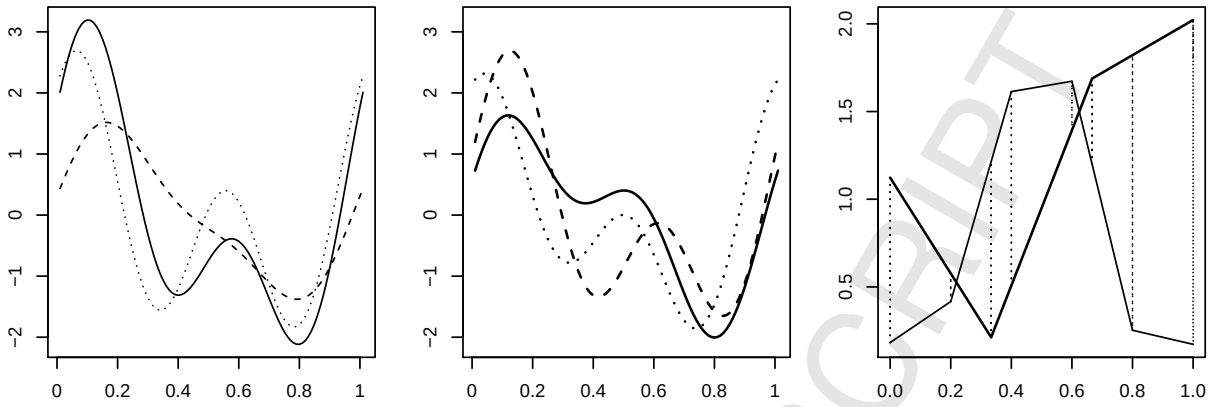

Figure 4: Left and mid panel: The solid curve has multiple intersections with each of the other two curves (left: Dimension 1, mid: Dimension 2). Right panel: For each dimension, the polygonal chains (thin: Dimension 1, bold: Dimension 2) show the chronological development of the intersection angles of the solid curves with the remaining curves. The distances between the polygonal chains at the support points are highlighted with dotted segments.

toy example is displayed in Figure 4. Assume a certain dependency structure across the dimensions for the majority of observations. Typically, multivariate outliers are assumed to have a different dependency structure, whereas marginal outliers are outlying w.r.t. one or more specific dimensions. Mostly we will be interested in outlying observations in which the dimensions are correlated to a lesser extent. For correlated dimensions, regardless if positive or negative, the area between the polygonal chains should be small. For rather uncorrelated dimensions, the area should be greater. This is due to the fact that if the dependency structure is similar in all dimensions, the intersection angle pattern will also be similar for each dimension. For computational ease, we use the mean of the absolute distances at the support points $\mathfrak{J} = \mathfrak{J}_1 \cup \mathfrak{J}_2$. Thereby we get a univariate measure for each function $i = 1, \ldots, n$, which we call SAPID, scaled average pairwise interangular distance. For notational simplicity, we state this measure in the case of two dimensions.

Definition 5. Scaled average pairwise interangular distance: SAPID

For $i = 1, \ldots, n$, define $\mathbf{x}_i$ and $\tilde{\mathbf{x}}$ as in Def. 3 having $p = 2$. Let $\gamma_{1,1}, \ldots, \gamma_{m_1,1}$ be the intersection angles that $\tilde{x}_1$ forms with the remainder of the functions in the first dimension. For the second dimension, the angles are named $\gamma_{1,2}, \ldots, \gamma_{m_2,2}$, respectively. Further, let $P_j = (P_{j,1}, ..., P_{j,m_j})$ be the simple polygonal chain formed by $P_{j,k} = \left(\frac{k-1}{m_j-1}, \gamma_{k,j}\right), k = 1, \ldots, m_j, j \in \{1, 2\}$. The scaled average pairwise interangular distance (SAPID) *is given by*

$$\text{SAPID}(\tilde{\mathbf{x}}) = |\mathfrak{J}|^{-1} \left(\sum_{k \in \mathfrak{J}_1} |\gamma_{k,1} - \gamma^*_{k,2}| + \sum_{k \in \mathfrak{J}_2} |\gamma^*_{k,1} - \gamma_{k,2}| - \sum_{k \in \mathfrak{J}_1 \cap \mathfrak{J}_2} |\gamma_{k,1} - \gamma_{k,2}| \right),$$

where $\gamma^*_{\cdot,1}$ contains the linear interpolations of P_1 at $\mathfrak{J}_2$ and $\gamma^*_{\cdot,2}$ contains the linear interpolations of P_2 at $\mathfrak{J}_1$.

This measure can also be generalised to p dimensions by means of the area or volume of simplices instead of absolute distances. Alternatively, univariate correlation measures for functional data were proposed in [25, 37].

We remark that SAPID should be applied to a sample which is free of marginal outliers. Marginal outliers can contaminate the vectors of intersection angles considerably, especially if they are not outlying in all dimensions. Then, in the first dimension one polygonal chain might have more intersections and more suspicious angles than in the second dimension, or vice versa. Finally, the actual outlier identification is conducted by means of a boxplot of the values of $\text{SAPID}(\mathbf{x}_1), \ldots, \text{SAPID}(\mathbf{x}_n)$. The upper outliers of this boxplot are taken to be the multivariate outliers of the dataset. A few properties of SAPID are mentioned in the subsequent remark.

Remark 5. *If two functions are identical in all dimensions, the corresponding* SAPID *values are also identical. Moreover, if all functions in a dataset are identical in all dimensions, the* SAPID *values become 0. This is of course*

also the lower bound of all possible SAPID *values. On the other hand, one can imagine an example similar to Fig. 3, where the curves have more than one dimension. Hence it can be obtained that the upper bound of* SAPID *values tends to* $\pi/2$.

2.2. Computational issues

The steps in the computation of FUNTA are sketched in Algorithm 1. Note that in line 2 in Alg. 1 the centring of the functions is described. In practice, the normed integration of a function can be approximated by the mean value of the function at the observed points. This of course accelerates computation. A further computational issue concerns the computation of intersection angles. When the data are available in high resolution, the intersection angles can be derived numerically. Surely, analytical derivatives would supply a slightly more accurate solution. On the other hand, numerical computation prevents intersection angles equal to 0 for non-identical functions. In the paragraph preceding Remark 2, this was shown to be possible for analytically derived angles.

Algorithm 1 Pseudo code of FUNTA

1: Fit differentiable functions $\mathbf{x}^* : [a, b] \to \mathbb{R}$ to the data
2: $\mathbf{x} = \mathbf{x}^* - \int \mathbf{x}^* d\mathbf{x}^* / |b - a|$
3: Derivatives: $\mathbf{x}'$ of $\mathbf{x}$
4: **for** $i = 1$ **to** n **do**
5: Remaining data: $\mathbf{x}_r = \mathbf{x}_{-i}$
6: **for** $j = 1$ **to** $n - 1$ **do**
7: Roots of $x_i - x_{r,j}$: $s_{i,j,1}, \ldots, s_{i,j,m_j}$
8: **for** $k = 1$ **to** m_j **do**
9: Intersection angle: $\gamma_{i,j,k} = \arccos\left(\frac{1+x_i'(s_{i,j,k})x_{r,j}'(s_{i,j,k})}{\sqrt{(1+x_i'(s_{i,j,k})^2)(1+x_{r,j}'(s_{i,j,k})^2)}}\right)$
10: **end for**
11: **end for**
12: FUNTA$(x_i) = 1 - \bar{\gamma}_{i,\cdot,\cdot}/\pi$
13: **end for**

The general idea behind FUNTA is similar to the Location-Slope depth [29], but disregards the location-specific part of the depth and is computationally sparser than a Location-Slope depth. For the multivariate or robustified versions of FUNTA, line 12 of the algorithm has to be adjusted appropriately.

2.3. Population versions of FUNTA pseudo-depth

Let X be a random function in the separable Hilbert space $\mathbb{H}$. Further let copies of X be $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{GP}(\mu(t), K(t_1, t_2))$ with μ the mean function and K the covariance function of the Gaussian process. From [1, Thm. 2.2.2] it follows that the derivatives of $X_i, i = 1, \ldots, n$ also exist and are defined as

$$X_i'(t) = \lim_{h \to 0} \frac{X_i(t + h) - X_i(t)}{h}.$$

Furthermore, $X_1', \ldots, X_n' \overset{\text{iid}}{\sim} \mathcal{GP}(\mu'(t), \frac{\partial^2 K}{\partial t_1 \partial t_2})$.

We assume that the number M of intersections of two curves is stochastic, but independent from the covariance function of the Gaussian process. Remember that we assume a preprocessing leading to centred data, e.g. forcing even highly positively correlated functions to intersect frequently following the independent errors. We also assume that the mean function is a constant function of time, i.e. $\mu(t) \equiv \mu$, therefore the intersection angles, which we denote by $\Gamma_1, \ldots, \Gamma_m$, are i.i.d. copies of a random variable Γ, hence we seek $\mathbb{E}(\Gamma)$.

Definition 6. *A population version of* FUNTA *pseudo-depth is given by*

$$\mathrm{FUNTA}(\tilde{X}) = 1 - (M\pi)^{-1} \sum_{k=1}^{M} \Gamma_i.$$

From Def. 6, it follows by the iterated expectation and Wald's equation that

$$\begin{aligned}
\mathbb{E}\left(\mathrm{FUNTA}\left(\tilde{X}\right)\right) &= 1 - \frac{1}{\pi}\mathbb{E}\left(M^{-1}\sum_{i=1}^{M}\Gamma_i\right) = 1 - \frac{1}{\pi}\mathbb{E}\left(\mathbb{E}\left(M^{-1}\sum_{i=1}^{M}\Gamma_i\middle| M\right)\right) \\
&= 1 - \frac{1}{\pi}\mathbb{E}\left(M^{-1}\mathbb{E}\left(\sum_{i=1}^{M}\Gamma_i\middle| M\right)\right) \overset{\text{iid}}{=} 1 - \frac{1}{\pi}\mathbb{E}\left(M^{-1}\times M\mathbb{E}\left(\Gamma\right)\right) \\
&= 1 - \frac{1}{\pi}\mathbb{E}\left(\arccos\left(\frac{1 + X_1' X_2'}{\sqrt{(1+(X_1')^2)(1+(X_2')^2)}}\right)\right).
\end{aligned}$$

Proposition 6. *Let $X(t) \sim \mathcal{GP}(\mu(t), K(t_1, t_2)), t_1, t_2 \in [a, b]$ with $\mu(t) \equiv \mu$ and $K(t_1, t_2) = f(|t_1 - t_2|), f : [a, b] \to \mathbb{R}_+$ continuous and twice continuously differentiable in 0. Then $X' \sim \mathcal{GP}(0, f^{(2)}(|t_1 - t_2|))$ and $X'(t) \sim \mathcal{N}(0, f^{(2)}(0))$.*

P . From [1, Thm. 2.2.2], it follows for the covariance function of X':

$$\begin{aligned}
\mu'(t) &= 0, \\
\frac{\partial K}{\partial t_1} &= \left.\frac{\partial f(x)}{\partial x}\right|_{|t_1 - t_2|} \frac{t_1 - t_2}{|t_1 - t_2|}, \\
\frac{\partial^2 K}{\partial t_1 \partial t_2} &= \left.\frac{\partial f(x)}{\partial x}\right|_{|t_1 - t_2|} \frac{-|t_1 - t_2| + (t_1 - t_2)\frac{t_1 - t_2}{|t_1 - t_2|}}{|t_1 - t_2|^2} \\
&\quad + \frac{t_1 - t_2}{|t_1 - t_2|} \left.\frac{\partial^2 f(x)}{\partial x^2}\right|_{|t_1 - t_2|} \frac{t_1 - t_2}{|t_1 - t_2|} \\
&= \left.\frac{\partial f(x)}{\partial x}\right|_{|t_1 - t_2|} \left[-\frac{|t_1 - t_2|}{|t_1 - t_2|^2} + \frac{(t_1 - t_2)^2}{|t_1 - t_2|^3}\right] \\
&\quad + \frac{(t_1 - t_2)^2}{|t_1 - t_2|^2} \left.\frac{\partial^2 f(x)}{\partial x^2}\right|_{|t_1 - t_2|} \\
&= f^{(2)}(|t_1 - t_2|).
\end{aligned}$$

Evaluation of the second parameter of the Gaussian process at $t_1 - t_2 = 0$ yields the variance of the univariate random variable $X'(t)$.

Corollary 1. *Let $X_1, X_2 \overset{iid}{\sim} \mathcal{GP}(\mu(t), K(t_1, t_2))$. Then $(X_1'(t), X_2'(t))^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0, 0)^T, \boldsymbol{\Sigma} = \mathrm{diag}(f^{(2)}(0), f^{(2)}(0)))$ $\forall\, t \in [a, b]$.*

2.4. Breakdown properties of FUNTA *pseudo-depth*

To underline the advantageous behaviour of robustified FUNTA against FUNTA, we look at the breakdown points of these estimators. Remember that we estimate the degree of shape outlyingness in a sample. Before we are able to derive breakdown points of FUNTA, we need to determine the maximum angle of an arbitrary function with others.

Lemma 1. *Denote* $g_c = \tilde{x}'(c)$ *with* $c = \underset{t}{\operatorname{argmax}}(|\tilde{x}'(t)|)$. *Then*

$$m^{-1} \sum_{i=1}^{n} \sum_{k=1}^{m_i} \gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})) \in [0, \gamma_{\sup}(\tilde{x}(c))) \tag{2}$$

with $\gamma_{\sup}(\tilde{x}(c)) := \arccos\left(-\sqrt{\frac{g_c^2}{1+g_c^2}}\right)$.

P . From $g_c = x'(c)$ and $c = \operatorname{argmax}_t(|\tilde{x}'(t)|)$ it follows that the maximal angle of $\tilde{x}$ and any function x_i from the sample can only occur in c. Let $g_c < 0$ and h_c be the descent of an arbitrarily extreme centred function which we assume w.l.o.g. to be x_1 in c with $x_1(c) = \tilde{x}(c)$. From $h_c \gg |g_c|$ we get $1 + g_c h_c < 0$. As we let go h_c to infinity, we get an upper bound for the maximal angle of any x_i with $\tilde{x}$:

$$\begin{aligned}
\cos(\gamma(\tilde{x}(c), \lim_{h_c\to\infty} x_1(c))) &= \lim_{h_c\to\infty} \frac{1 + g_c h_c}{\sqrt{1+g_c^2}\sqrt{1+h_c^2}} \\
&= -\frac{1}{\sqrt{1+g_c^2}} \lim_{h_c\to\infty} \left| \left(\frac{(1+g_c h_c)^2}{1+h_c^2} \right)^{1/2} \right| \\
&= -\frac{1}{\sqrt{1+g_c^2}} \lim_{h_c\to\infty} \left| \left(\underbrace{\frac{1}{1+h_c^2}}_{\to 0} + \underbrace{\frac{2g_c h_c}{1+h_c^2}}_{\to 0} + \underbrace{\frac{g_c^2 h_c^2}{1+h_c^2}}_{\to g_c^2} \right)^{1/2} \right| \\
&= -\sqrt{\frac{g_c^2}{1+g_c^2}}.
\end{aligned}$$

Thereby it holds that $\gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})) < \arccos\left(-\sqrt{\frac{g_c^2}{1+g_c^2}}\right)$ from which Eq. (2) follows. For $g_c \geq 0$ and $h_c \to -\infty$, the steps are similar and the result is the same.

Corollary 2. *In the situation of Lemma 1,* $\mathrm{FUNTA}_n(\tilde{x}) \in (1 - \gamma_{\sup}(\tilde{x}(c))/\pi, 1]$.

We consider finite-sample breakdown points in the sense of [11] as a measure of robustness. With respect to our pseudo-depth the breakdown point can then be defined as the smallest proportion of replaced functions in the sample $\mathbf{x} = (x_1, \ldots, x_n)$ which can send the value of $\mathrm{FUNTA}_n(\tilde{x})$ to the boundary of the parameter space, which is given by Corollary 2. As in the case of scatter estimators, we distinguish between implosion and explosion breakdown points. In contrast to usual scatter estimators, the bounds of the domain are not $[0, \infty)$. Therefore we define the finite-sample implosion breakdown point as the maximum fraction of contaminated observations that does not shift FUNTA to 1. For a more formal exposition assume that $\mathbf{x}^\varepsilon$ denotes a sample where k functions of $\mathbf{x}$ are replaced by arbitrary functions, $\varepsilon = k/n$. The finite-sample implosion breakdown point is then given by

$$\varepsilon^*_{imp}(\mathbf{x}, \mathrm{FUNTA}_n(\tilde{x})) = \min\{\varepsilon : \sup_{\mathbf{x}^\varepsilon} \mathrm{FUNTA}_n(\tilde{x}) = 1\}.$$

The more interesting case occurs when a replacement of functions evokes extreme angles. The finite-sample explosion breakdown point is given by

$$\varepsilon^*_{exp}(\mathbf{x}, \mathrm{FUNTA}_n(\tilde{x})) = \min\{\varepsilon : \inf_{\mathbf{x}^\varepsilon} \mathrm{FUNTA}_n(\tilde{x}) = 1 - \gamma_{\sup}(\tilde{x}(c))/\pi\}.$$

However, the maximum intersection angle of one function with another does not give an immediate solution to the explosion breakdown point problem. It is obvious that if the contaminated function has many intersections with a regular function, this influences the average angle more than if it has only one intersection. The problem is that the size of the other intersection angles is only known under certain conditions:

Example 1. *Let* $\tilde{x}'(t) \equiv \tilde{c}, x_1'(t) \equiv c_1, \ldots, x_n'(t) \equiv c_n$ *with* $\tilde{c}, c_j > 0 \; \forall \; j = 1, \ldots, n$. *Then replacing w.l.o.g.* $x_1(t)$ *by a multivariate sequence of functions* $\check{x}_1^k(t) = \tilde{x}(t) + r^k \sin(r^k t)$ *causes the number of intersections of* $\tilde{x}$ *with* $\check{x}_1$ *to tend to infinity for* $r^k \overset{k\to\infty}{\longrightarrow} \infty$. *Half of the intersection angles is given by* $\gamma_{\sup}(\tilde{x}(c))$, *the other half is given by* $\pi - \gamma_{\sup}(\tilde{x}(c))$, *such that the average angle approaches* $\pi/2$. *Therefore we get* $\text{FUNTA}_n(x_1) = 1/2$. *Since* $\tilde{x}$ *has only a finite number of intersections with every other centred function* $x_2, \ldots, x_n$, *the finite-sample explosion breakdown point of* FUNTA *of* $\tilde{x}$ *is* $1/n$. *However, if we drop the assumption of every function having a positive slope, it is possible that the mean intersection angle* before *manipulating the sample is larger than* $1/2$. *This shows that the finite-sample explosion breakdown point highly depends on the given sample* $\mathbf{x}$ *such that general results are difficult to derive.*

Proposition 7. *Let* $x_1(t), \ldots, x_n(t), t \in [a, b]$. *The finite-sample implosion breakdown point of* $\text{FUNTA}_n(\tilde{x})$ *is* $1/n$.

P . Replace w.l.o.g. x_2 with $\check{x}_2$, which is tangent to $\tilde{x}$ in m_2 equidistant points $a < t_1 < \ldots < t_{m_2} < b$. Therefore we have m_2 intersection angles $\gamma_1 = \ldots = \gamma_{m_2} = 0$. Then we get $\lim_{m_2\to\infty} m^{-1} \sum_{i=1}^{n} \sum_{k=1}^{m_i} \gamma_k(\tilde{x}(s_{i,k}), x_i(s_{i,k})) = 0$ and therefore $\text{FUNTA}_n(\tilde{x}) = 1$.

Proposition 8. *Let* $x_1(t), \ldots, x_n(t), t \in [a, b]$. *The finite-sample implosion and explosion breakdown point of* $\text{rFUNTA}_n(\tilde{x})$ *is given by* $\lfloor \frac{n+1}{2} \rfloor / n$.

P . This is due to the fact that the median of the maximum angles is taken. Thus rFUNTA inherits the finite-sample breakdown point over the maximal angles from the median. Manipulating a maximum angle for the implosion breakdown point corresponds to the exchange of one element in the sample as described in the proof of Prop. 7, but here we only need exactly one intersection with intersection angle 0 for each of the $\lfloor \frac{n+1}{2} \rfloor$ contaminated observations to shift $\text{rFUNTA}_n(\tilde{x})$ to 1. An explosion of robustified FUNTA can only occur if the maximum intersection angle of $\tilde{x}$ from Lemma 1 is returned by the median. This happens for example when the contaminated functions are identical and only have one intersection with $\tilde{x}$ in c. Therefore at least $\lfloor \frac{n+1}{2} \rfloor$ functions manipulated in the style of $\check{x}_2$ are needed to shift $\text{rFUNTA}_n(\tilde{x})$ arbitrarily close to the lower interval bound of Corollary 2.

3. A simulation study for outlier detection

As the distribution of FUNTA is not completely known, model-based outlier detection techniques like the concept of α-outliers [10] are not directly applicable. The same problem exists for other functional depth measures [12]. A smoothed bootstrap procedure based on weighting runs was introduced by [12]. The asymptotic validity of this approach has not been proved yet [20]. Though, it yields sensible results. By interpreting the cutoff value from the bootstrap procedure as estimated quantile q of the distribution of the functional depths, we find the link to the concept of α-outliers: Under mild assumptions w.r.t. the distribution, the α-outlier region is equivalent to the lower α-tail region [24, Lemma 6.1(c)].

In this simulation study we compare the performance of FUNTA and rFUNTA with some well-known functional depths: h-modal depth [9, 13, HMD], modified band depth [27, 34, MBD] and the integrated depth [16, FMD]. For h-modal depth, we tested each of the combinations of metrics (L_1, L_2, L_∞) and trimming percentages $(tr = 0.01, tr = 0.25)$. However, the variation of the threshold value was too small to change any set of outliers. Modified band depth was computed using $J = 2$. Generally, we perform the simulations on the centred data as well as on the uncentred data. We expect that for shape outliers it is useful to centre the data in advance and FUNTA pseudo-depths only work for centred data. The performance of the outlier detection is measured by the number of outliers correctly identified as outliers (true positives, TP) as well as the number of regular curves falsely identified as outliers (false positives, FP).

3.1. Data situation 1: Smooth data

In data situation 1 we generate a sample of $n = 110$ functions $x_1(t), \ldots, x_{110}(t)$ from a Fourier basis

$$x_i(t) = y_{i,1} + \sqrt{2} \sum_{\ell=1}^{4} (y_{i,2\ell} \sin(\ell\omega t) + y_{i,2\ell+1} \cos(\ell\omega t)), \quad i = 1, \ldots, 110, t \in [0, 1],$$

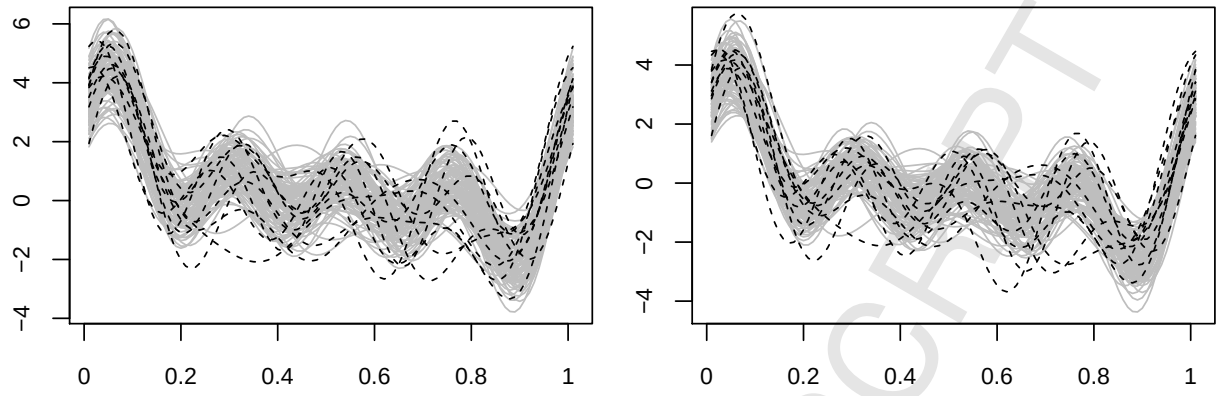

Figure 5: Data situation 1. Left panel: Uncentred. Right panel: Centred. The regular curves are solid grey, the outlying curves are dashed black.

where $\omega = 2\pi$. The entries of $\mathbf{y}_i \in \mathbb{R}^9, i = 1, \ldots, 100$ are drawn from a Beta$(2, 2)$ distribution and the entries of $\mathbf{y}_i \in \mathbb{R}^9, i = 101, \ldots, 110$ from a Beta$(0.5, 0.5)$ distribution. Therefore the first 100 functions can be seen as regular observations and the last 10 functions as outliers. The simulated data are depicted in Fig. 5.

Table 1: Performance of depth measures for data situation 1. Outlier identification based on centred data is indicated by "c", uncentred data by "u". The left part of the table shows the results for the contaminated data, the right part shows them for the outlier-free data.

	With outliers						Without outliers		
q	0.01		0.05		0.1		0.01	0.05	0.1
Depth	TP	FP	TP	FP	TP	FP	FP	FP	FP
FUNTA, c	1	0	6	0	7	2	0	3	8
rFUNTA, c	0	0	5	0	8	1	1	4	6
HMD, c	1	0	9	0	10	1	0	2	6
MBD, c	2	0	5	1	7	5	0	4	9
FMD, c	2	0	4	2	7	5	1	4	9
HMD, u	1	0	10	0	10	1	0	2	8
MBD, u	1	0	6	1	9	3	0	3	6
FMD, u	1	0	5	1	7	4	1	5	8

The number of detected outliers is summarised in Table 1. We see that h-modal depth clearly outperforms every other depth in the study, both for each uncentred and centred data. This is not very surprising owing to the fact that the data is nearly centred beforehand. We also see that the integrated depth and modified band depth perform quite similar, with slight advantages for modified band depth. Furthermore FUNTA performs slightly better for $q \leq 0.05$ whereas rFUNTA shows better results for $q = 0.1$. However both FUNTA pseudo-depths are quite similar to modified band depth and integrated depth, although it is obvious that FUNTA pseudo-depths identify less false positives for $q = 0.1$. For the data without additional outliers (100 observations), h-modal depth of the centred data still outperforms the other depths. For the uncentred data, the results are not so clear—modified band depth seems to be a reasonable choice here, too.

3.2. Data situation 2: Noisy data

Using R [32] function `sde.sim` [23], we generate 100 stochastic processes with initial value $X_0 = 20$, and drift coefficient $\log(x)$, diffusion coefficient $\sigma_1 = 1.5$ and $T = 101$ as regular observations. The outliers are 10 stochastic processes with the coefficients mentioned above but with increased $\sigma_2 = 3$. The processes satisfy the stochastic differential equation

$$dX_t = \log(X_t)dt + \sigma_i dW_t, \quad i = 1, 2$$

with W_t as Wiener process. The simulated data are shown in Fig. 6.

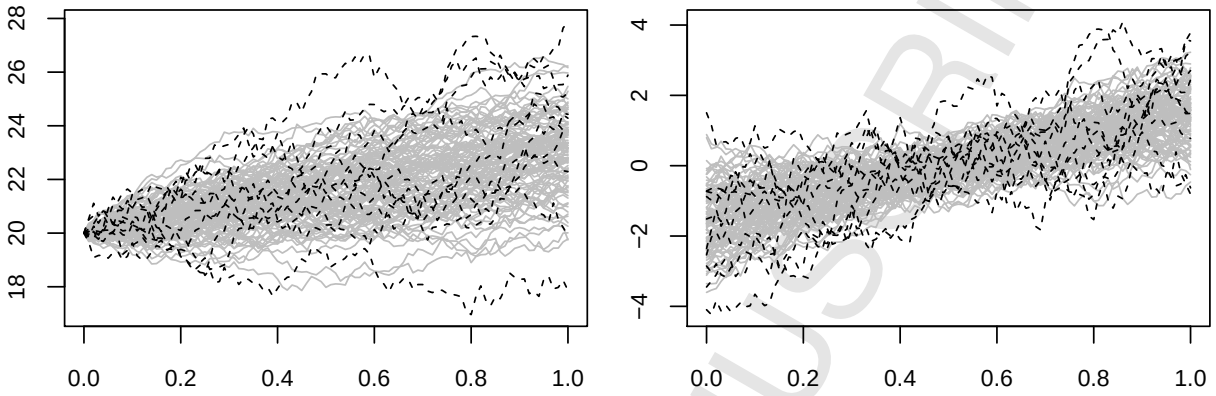

Figure 6: Data situation 2. Left panel: Uncentred. Right panel: Centred. The regular curves are solid grey, the outlying curves are dashed black.

Table 2: Performance of depth measures for data situation 2. Outlier identification based on centred data is indicated by "c", uncentred data by "u". The left part of the table shows the results for the contaminated data, the right part shows them for the outlier-free data.

	With outliers						Without outliers		
q	0.01		0.05		0.1		0.01	0.05	0.1
Depth	TP	FP	TP	FP	TP	FP	FP	FP	FP
FUNTA, c	1	0	6	0	10	0	0	2	3
rFUNTA, c	1	0	7	0	10	0	0	2	3
HMD, c	4	1	6	2	7	10	1	5	15
MBD, c	1	1	4	3	4	9	1	5	15
FMD, c	1	0	4	2	4	7	1	5	12
HMD, u	2	2	4	6	5	9	2	6	11
MBD, u	1	2	2	4	3	10	3	6	11
FMD, u	1	0	2	3	3	10	1	5	12

For such wiggly data FUNTA pseudo-depths seem to be very powerful, especially the robustified version. Looking at true and false positives again, see Table 2, h-modal, modified band and integrated depth perform better on the centred dataset than on the uncentred, particularly for $q = 0.05$. While using the three classic functional depths leads to many false positives, FUNTA pseudo-depths do not detect a single wrong outlier. Among the classic depths, h-modal depth outperforms the others w.r.t. the identification of true positives. When we consider the "in control" situation of a dataset, the results from the second part of Table 2 clearly suggest to choose either of the FUNTA pseudo-depths. The other depths perform astoundingly similar. For such kind of noisy data we recommend to choose rFUNTA.

3.3. Bivariate noisy data situations

To check whether FUNTA's ability to detect outliers in wiggly data still holds in the bivariate case, we generate a couple of data situations. In each situation we have two dimensions where the data for each dimension are generated according to data situation 2. These data situations are based on Sec. 3.2 and can be summarised as follows: For both dimensions of the next three data situations, we generate 100 regular observations using the `sde` package and

$\sigma_1 = 1.5$. For data situation 3, we set $\sigma_2 = 3$ in dimension 1 and $\sigma_2 = 1.5$ in dimension 2, therefore we have extreme outliers in one dimension and no outliers in the other dimension. For data situation 4, we choose $\sigma_2 = 3$ in both dimensions and get 10 extreme outliers in both dimension. For data situation 5, moderate outliers are generated using $\sigma_2 = 2$ in both dimensions. As a competing depth we choose the multivariate functional halfspace depth [7, MFHD] using the R package `MFHD` [22].

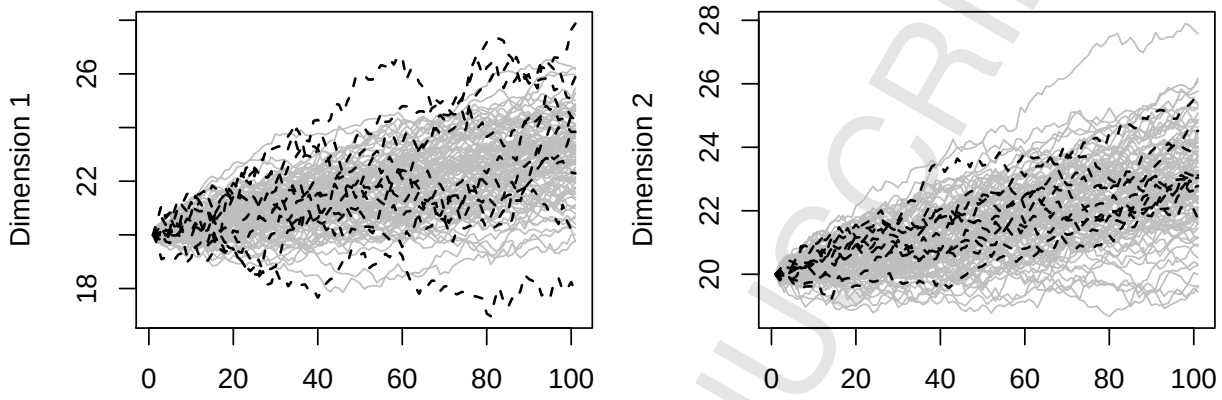

Figure 7: Data situation 3 (uncentred). The regular curves are solid grey. The dashed black curves on the left panel are outliers, those on the right panel are the corresponding regular observations.

In this section, data situations without outliers (i.e., $\sigma_2 = 1.5$) are identical and we state the number of false positives in Table 6. We remark that for the smoothed bootstrap procedure it is necessary to estimate the covariance matrices separately for each dimension.

The generated data for data situation 3 is shown in Fig. 7. In Tab. 3 we see that FUNTA and rFUNTA lose only little of their sensitivity. For rFUNTA this is less surprising than for FUNTA, since the mean intersection angle of an outlying function should decrease considerably when only half of the dimensions are contaminated. The maximum intersection angle of rFUNTA will not change in most cases, but the computation of the threshold will be affected by the added uncontaminated dimension. The MFHD behaves worse than FUNTA pseudo-depths, furthermore the centring of the data does not seem to improve the results. Perfect results are obtained for the combination of SAPID and FUNTA respectively rFUNTA. This shows a misspecified contamination level q can be compensated by SAPID.

Table 3: Performance of depth measures for data situation 3.

	Uncentred						Centred					
	$q = 0.01$		$q = 0.05$		$q = 0.1$		$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
FUNTA	-	-	-	-	-	-	1	0	5	0	10	0
FUNTA & SAPID	-	-	-	-	-	-	10	0	10	0	10	0
rFUNTA	-	-	-	-	-	-	1	0	7	0	8	0
rFUNTA & SAPID	-	-	-	-	-	-	10	0	10	0	10	2
MFHD	2	2	2	7	3	11	0	0	3	6	5	12

When both dimensions are contaminated on the same observations, we expect that the performance of the algorithms improves. We find in Tab. 4 that this is barely the case. FUNTA pseudo-depths remain more or less unchanged, whereas MFHD reveals some improvements w.r.t. true positives as well as false positives. Still, FUNTA and rFUNTA clearly outperform MFHD. SAPID still detects all remaining outliers, but also some false positives. However SAPID clearly outperforms the one-step procedures for $q < 0.1$.

For the results concerning moderate outliers in both dimensions (Fig. 8) we look at Tab. 5. Again, the results

Table 4: Performance of depth measures for data situation 4.

	Uncentred						Centred					
	$q = 0.01$		$q = 0.05$		$q = 0.1$		$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
FUNTA	-	-	-	-	-	-	1	0	5	0	10	0
FUNTA & SAPID	-	-	-	-	-	-	10	0	10	1	10	2
rFUNTA	-	-	-	-	-	-	1	0	6	0	10	0
rFUNTA & SAPID	-	-	-	-	-	-	10	0	10	2	10	2
MFHD	3	1	4	4	5	7	2	2	4	7	5	9

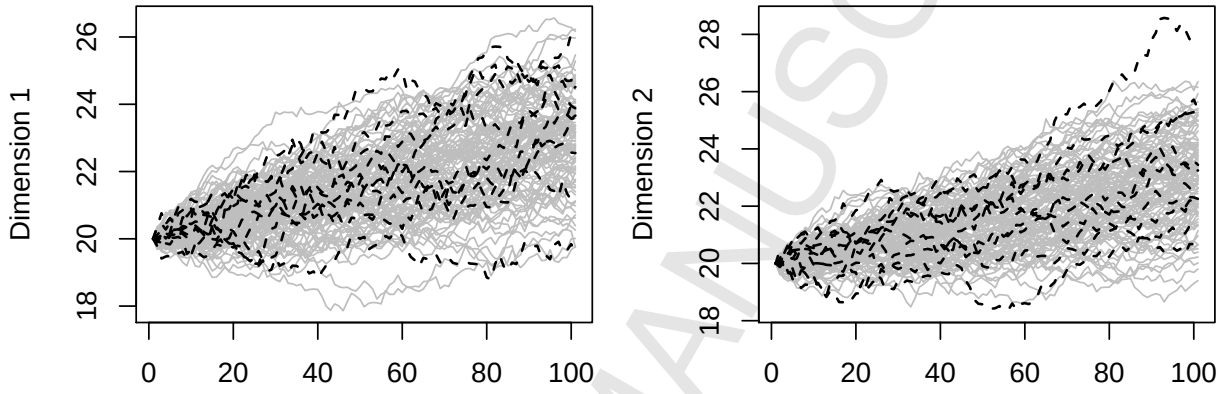

Figure 8: Data situation 5 (uncentred). Left panel: Dimension 1. Right panel: Dimension 2. The regular curves are solid grey. The dashed black curves are outliers.

for MFHD are worse than those of FUNTA pseudo-depths. Particularly for $q = 0.05$ rFUNTA is clearly better than FUNTA. For $q = 0.1$ rFUNTA yields the perfect result of every observation classified correctly. We point out that the loss of sensitivity for both FUNTA pseudo-depths—compared to the extreme outlier situation 4—is relatively small. In contrast, the results of MFHD compared to data situation 4 are worse in nearly every component. In contrast to the previous two data situations, now SAPID does not make the detection rate perfect. For small q, this is still a sensible extension of FUNTA and rFUNTA and for large q, SAPID does not cause to many false positives, especially when compared to MFHD.

Table 5: Performance of depth measures for data situation 5.

	Uncentred						Centred					
	$q = 0.01$		$q = 0.05$		$q = 0.1$		$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP	TP	FP
FUNTA	-	-	-	-	-	-	1	0	2	0	9	0
FUNTA & SAPID	-	-	-	-	-	-	8	2	9	2	10	2
rFUNTA	-	-	-	-	-	-	0	0	5	0	10	0
rFUNTA & SAPID	-	-	-	-	-	-	8	1	9	1	10	2
MFHD	0	3	3	7	4	13	0	2	2	7	2	11

The results of Tab. 6 are similar to the right part of Tab. 2. rFUNTA even falsely detects two observations as outliers less than in the corresponding univariate situation. FUNTA also performs very well. MFHD detects more false outliers than expected ($100q$) and is therefore outperformed by the other depths. The two-step procedure of

SAPID and FUNTA or rFUNTA respectively, has a few false positives more, but is still in the region of the expected number of outliers.

Table 6: Performance of depth measures for the wiggly bivariate data situations without outliers. Entries are number of regular curves identified as outliers.

	Uncentred			Centred		
	$q = 0.01$	$q = 0.05$	$q = 0.1$	$q = 0.01$	$q = 0.05$	$q = 0.1$
FUNTA	-	-	-	0	1	3
FUNTA & SAPID	-	-	-	2	4	4
rFUNTA	-	-	-	0	0	3
rFUNTA & SAPID	-	-	-	2	2	3
MFHD	2	7	14	2	9	11

We also generate a data situation in which the dependency structure across the dimensions separates regular data from outliers. For $t = 1, \ldots, T$, let x_t be a k-dimensional time series. A p-th order vector autoregression ($VAR(p)$ model, see [17]) is defined as

$$x_t = c + \sum_{i=1}^{p} \mathbf{\Phi}_i x_{t-1} + \varepsilon_t,$$

where $c \in \mathbb{R}^k$, $\mathbf{\Phi}_i \in \mathbb{R}^{k \times k}, i = 1, \ldots, p$, $\mathbb{E}(\varepsilon_t) = 0, \mathbb{E}(\varepsilon_t \varepsilon_\tau^T) = \mathbf{\Sigma}$ for $t = \tau$ and $\mathbb{E}(\varepsilon_t \varepsilon_\tau^T) = 0$ for $t \neq \tau$. The data generation has been conducted using the R package `MTS` [36] and the parameters $p = 1, k = 2, c = 0, \mathbf{\Phi}_1 = \begin{pmatrix} 0.8 & 0.3 \\ -0.4 & 1 \end{pmatrix}$. For the first 100 regular observations we used $\mathbf{\Sigma}_1 = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$ and for the outliers we used $\mathbf{\Sigma}_2 = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}$. Hence we obtain bivariate observations with different dependency structures across the dimensions. Consult Fig. 9 for some examples of the generated curves. We assume x_t as the discrete realisations of an underlying smooth time-continuous process.

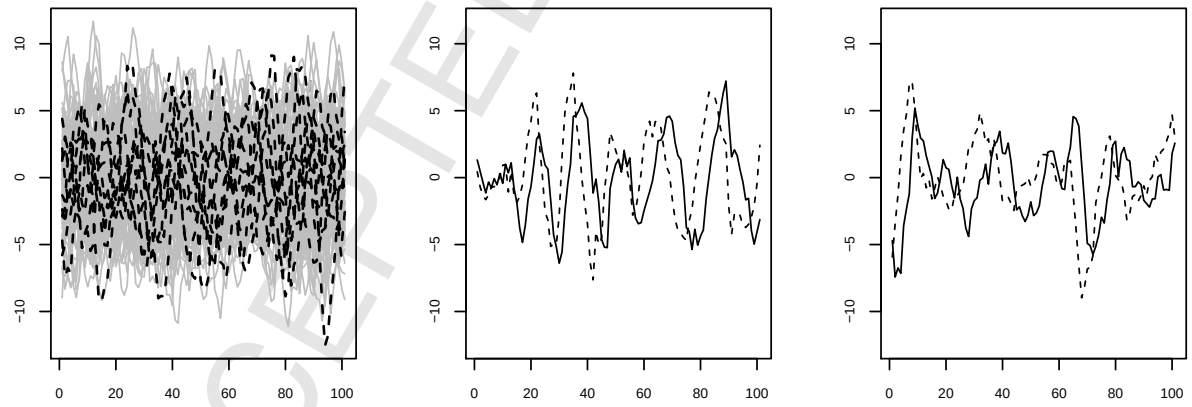

Figure 9: Data situation 6 (uncentred). Left panel: All observations in dimension 1. We omitted a plot of the observations in dimension 2 since they are very similar. Mid panel: Dimension 1 (solid) and 2 (dashed) of a regular observation. The dimensions look quite similar, but seem slightly shifted. Right panel: Dimension 1 (solid) and 2 (dashed) of an outlier. The correlation across the dimensions seems to be smaller.

The results shown in Table 7 exhibit that the combination of FUNTA or rFUNTA and SAPID are comparable to MFHD in most cases. The number of true positives is always between 1 and 3, as for the MFHD. However, the proposed methods detect less false positives than MFHD. The differences for the results of MFHD between centred

Table 7: Performance of depth measures for VAR(1) model.

	$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP
FUNTA & SAPID, c	2	1	2	3	3	10
FUNTA, c	0	0	2	3	3	8
rFUNTA & SAPID, c	2	1	3	1	1	3
rFUNTA, c	0	0	1	1	1	3
MFHD, c	1	1	2	6	3	13
MFHD, u	1	1	3	6	3	12

and uncentred data are very small. Since FUNTA and rFUNTA are quite conservative for small q, we observe the largest discrepancy between the results obtained with or without SAPID for $q = 0.01$. The best result in the table is gained for the combination of rFUNTA and SAPID and $q = 0.05$. However, the intensity of the improvement caused by the additional SAPID is not monotone with respect to q. The deletion of false positives detected by FUNTA or rFUNTA can lead to the detection of further false positives by SAPID (first row, last entry) or even to the non-identification of outliers which have been detected for a smaller q (third row, fifth entry).

4. Real data example

4.1. Univariate real data example

Consider the phoneme data example [18, 19], an experiment for speech recognition. For this dataset the voices of 50 male speakers were recorded. Each speaker had to read aloud the same word list, and the recordings were classified into five different phonemes ("aa", "ao", "sh", "dcl", "iy"). We used a subset of this dataset [14, see Chap. 8] and drew 100 consecutive observations of the "ao" phoneme and 10 consecutive observations of the "aa" phoneme. We aim at detecting the "aa" phonemes. After centring, the curves are hard to distinguish (see Fig. 10). Particularly, in the first quarter of the data no structural difference can be detected visually. In the other quarters a slight pattern of the "aa" phonemes might be anticipated, but it could also arise from the smaller sample size of this phoneme.

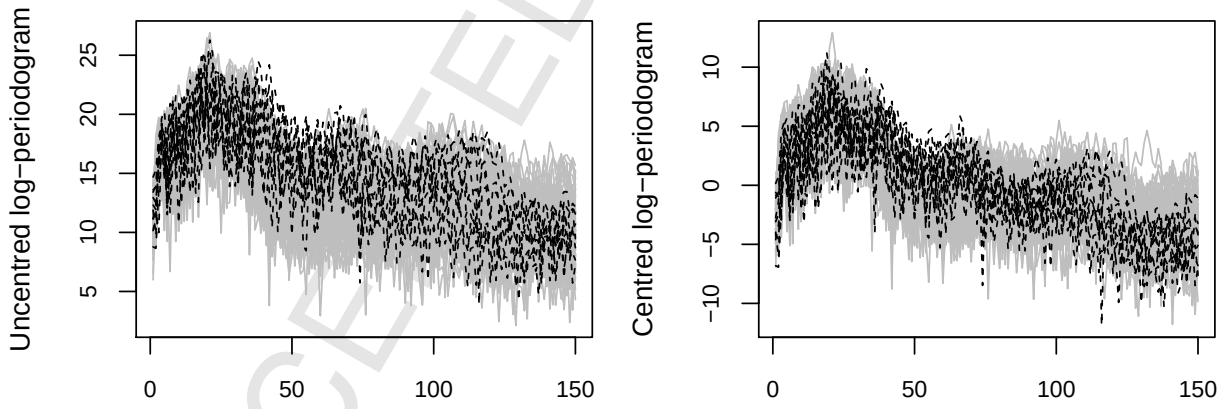

Figure 10: Excerpt from the Phoneme data example. Solid grey curves: "ao"; dashed black curves: "aa". Left panel: Uncentred data. Right panel: Centred data. The panels are similar.

Using the same smoothed bootstrap procedure as in Sec. 3, we conduct outlier detection for the phoneme dataset. Table 8 shows the performance of the depth measures. For very small q modified band depth and integrated depth have the best performance among the depths. But as q becomes larger, the performance of these depths decrease more and more. For medium q, FUNTA works best for the data, for large q, h-modal depth and FUNTA outperform the

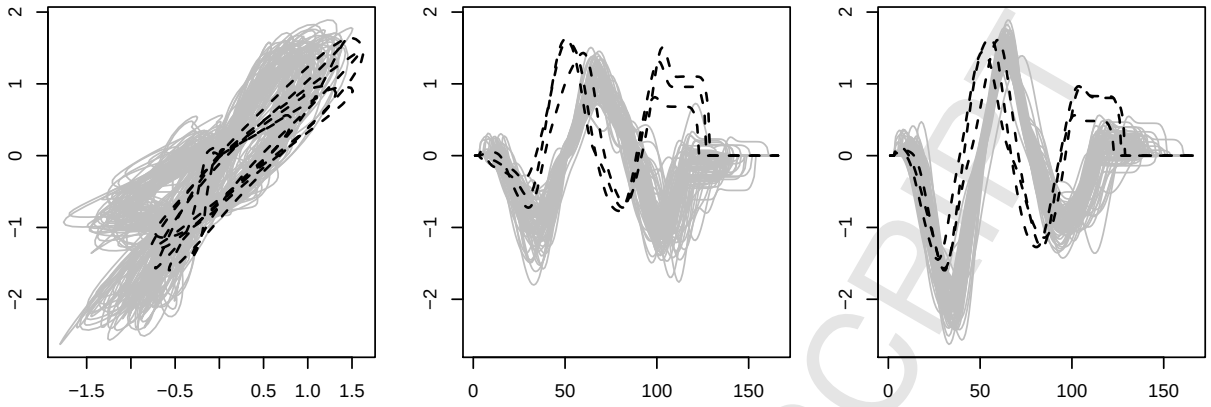

Figure 11: Left panel: The velocity trajectories of "b" (solid, grey) and "u" (dashed, black) in $\mathbb{R}^2$. Mid panel: The velocity in the x-direction over time. Right panel: The velocity in the y-direction over time.

other depths, whereupon FUNTA is slightly more conservative. The robustified FUNTA is outperformed for $q < 0.1$. Only for $q = 0.1$ it does not detect the least outliers.

Table 8: Performance of depth measures for centred phoneme data.

	$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP
FUNTA	1	0	3	2	5	7
rFUNTA	0	1	2	5	4	8
HMD	1	0	3	3	6	8
MBD	2	0	2	2	3	8
FMD	2	0	2	4	3	8

4.2. Bivariate real data example

We examine a dataset concerning pen tip trajectories for handwritten letters, obtained from [26]. For each lowercase letter that can be written without lifting the pen, approximately 100 three-dimensional pen tip trajectories were recorded and numerically differentiated. A more exhaustive explanation of the dataset is given in [38].

The velocity of the pen in the x-direction forms the first dimension of our dataset and the velocity in the y-direction forms the second dimension. We restrict to the 73 trajectories belonging to the letter "b", but add three trajectories of the letter "u" to check whether these can be identified as outliers.

The results shown in Tab. 9 are quite similar for the measures except for the first two rows. We see that all competitors detect the three outliers sooner or later. SAPID improves the detection of outliers for small q, but also leads to a few more false positives. The best result is found for rFUNTA and $q = 0.05$, where all outliers and only one false positive is found. MFHD performs only slightly worse w.r.t. the centred data. One interesting issue arises when we compare FUNTA for $q = 0.05$ and $q = 0.1$. While the outliers are already found for the smaller q, the larger q leads to a strong increase in false positives. Here, rFUNTA clearly outperforms FUNTA.

5. Conclusions and outlook

In this paper we introduced two versions of an angle-based multivariate functional measure, named FUNTA and rFUNTA. These pseudo-depths are specifically designed to detect shape outliers in functional data. We showed that FUNTA has a finite-sample implosion breakdown point equal to $1/n$ and that the finite-sample implosion and

Table 9: Performance of depth measures for the character trajectories dataset.

	$q = 0.01$		$q = 0.05$		$q = 0.1$	
	TP	FP	TP	FP	TP	FP
FUNTA	1	0	3	2	3	11
FUNTA & SAPID	3	5	3	5	3	12
rFUNTA	1	0	3	1	3	4
rFUNTA & SAPID	3	5	3	5	3	7
MFHD, u	1	0	2	3	2	5
MFHD, c	0	0	3	2	3	5

explosion breakdown points of rFUNTA equal the finite-sample breakdown point of the univariate median. Moreover a population version of FUNTA was derived based on the concept of Gaussian processes. One advantage of FUNTA can be seen in the needlessness to choose tuning parameters, unlike other functional depths like h-modal or modified band depth. For the detection of outliers, we of course need to specify the quantile q in the bootstrap procedure for all depths, including FUNTA and rFUNTA. In a simulation study and a real data example we found that FUNTA pseudo-depths are comparable to classical functional depths or, in some cases, even superior. Since the simulation results of FUNTA pseudo-depths were promising, this might encourage further work on complexity reduction in multi-functional data.

The multivariate extensions of FUNTA pseudo-depths are examined in the simulation studies and combined with a new measure named SAPID which focuses on the discrepancy of the intersection angle patterns across the dimensions. We have seen that the FUNTA pseudo-depths mostly improve when, as a second step, outlier detection using SAPID is performed. An increased identification of false positives was rather not the case, especially when the correct marginal outliers have been removed beforehand. Similar results can be found for the bivariate real data example.

As always in functional data, the amount of smoothing has to be chosen carefully. If the data are not smooth, the practitioner has to decide whether noise is a general problem of the process or measurement technique or if it may incorporate important information for specific curves. In particular, it has turned out in additional simulations that oversmoothing had a considerably smaller effect on the performance of FUNTA than undersmoothing. This is due to the fact that undersmoothing causes additional intersections in the data and therefore non-informative angles contaminate the average intersection angle. Another pre-processing technique in functional data analysis is time warping. In our opinion, time warping prior to outlier detection transforms outliers into regular observations. Therefore, the interplay of time warping and FUNTA is not examined in this paper.

In linear models, high leverage points indicate observations that influence the estimated coefficients of the model much more than the remaining observations. Analogously, it might be interesting to examine how much influence one functional observation has on the FUNTA values of the other observations. This could also help to detect outliers using a suitable identification rule. Multi-step procedures for univariate functional outlier detection based on FUNTA were beyond the scope of this paper, but could be particularly interesting because FUNTA and rFUNTA are quite conservative w.r.t. outlier detection. Further investigation of the properties of SAPID will be needed. Since the bootstrap procedure is computationally demanding, it is an important question how to make the threshold detection more user-friendly. Applications in random fields might be a natural extension of FUNTA, but the definition or computation of intersection angles in $\mathbb{R}^3$ or higher is not straightforward.

Acknowledgements. The financial support of the Deutsche Forschungsgemeinschaft (SFB 823, project B1) is gratefully acknowledged. We thank the anonymous reviewer and the editors for their valuable remarks.

References

[1] R.J. Adler, 1981. The Geometry of Random Fields. Wiley, Chichester.
[2] G. Aneiros, P. Vieu, 2015. Partial linear modelling with multi-functional covariates. Computation. Stat. doi:10.1007/s00180-015-0568-8.
[3] A. Arribas-Gil, J. Romo, 2014. Shape outlier detection and visualization for functional data: the outliergram. Biostatistics 15, 603–619.
[4] V. Barnett, T. Lewis, 1994. Outliers in Statistical Data. 3rd ed., Wiley, New York.

[5] R. Biasi, F. Ieva, A.M. Paganoni, N. Tarabelloni, 2014. Multivariate functional data depth measure based on variance-covariance operators, in: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna, pp. 49–54.

[6] E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), 2014. Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna.

[7] G. Claeskens, M. Hubert, L. Slaets, K. Vakili, 2014. Multivariate functional halfspace depth. J. Am. Stat. Assoc. 109, 411–423.

[8] J.A.A. Collazos, R. Dias, A.Z. Zambom, 2015. Consistent variable selection for functional regression models. J. Multivariate Anal. doi:10.1016/j.jmva.2015.06.007.

[9] A. Cuevas, M. Febrero, R. Fraiman, 2007. Robust estimation and classification for functional data via projection-based depth notions. Computation. Stat. 22, 481–496.

[10] P.L. Davies, U. Gather, 1993. The identification of multiple outliers. J. Am. Stat. Assoc. 88, 784–792.

[11] D.L. Donoho, P.J. Huber, 1983. The notion of breakdown point, in: P.J. Bickel, K.A. Doksum, J.L. Hodges (Eds.), A Festschrift for Erich L. Lehmann. Wadsworth International Group, Belmont, Calif.. Wadsworth statistics/probability series, pp. 157–184.

[12] M. Febrero, P. Galeano, W. González-Manteiga, 2008. Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels. Environmetrics 19, 331–345.

[13] M. Febrero-Bande, M. Oviedo de la Fuente, 2012. Statistical computing in functional data analysis: The R package fda.usc. J. Stat. Softw. 51, 1–28. URL: http://www.jstatsoft.org/v51/i04/.

[14] F. Ferraty, P. Vieu, 2006. Nonparametric Functional Data Analysis: Theory and Practice. Springer series in statistics, Springer, New York.

[15] F. Ferraty, P. Vieu, 2009. Additive prediction and boosting for functional data. Comput. Stat. Data An. 53, 1400–1413.

[16] R. Fraiman, G. Muniz, 2001. Trimmed means for functional data. Test 10, 419–440.

[17] J.D. Hamilton, 1994. Time Series Analysis. Princeton University Press, Princeton.

[18] T.J. Hastie, A. Buja, R. Tibshirani, 1995. Penalized discriminant analysis. Ann. Statist. 23, 73–102.

[19] T.J. Hastie, R. Tibshirani, J.H. Friedman, 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer series in statistics. 2nd ed., Springer, New York.

[20] L. Horváth, P. Kokoszka, 2012. Inference for Functional Data with Applications. Springer series in statistics, Springer, New York.

[21] M. Hubert, P.J. Rousseeuw, P. Segaert, 2015. Multivariate functional outlier detection. Stat. Methods Appl. doi:10.1007/s10260-015-0297-8.

[22] M. Hubert, K. Vakili, 2013. MFHD: Multivariate functional halfspace depth. URL: http://CRAN.R-project.org/package=MFHD. R package version 0.0.1.

[23] S.M. Iacus, 2008. Simulation and Inference for Stochastic Differential Equations: With R Examples. Springer series in statistics, Springer, New York.

[24] S. Kuhnt, A. Rehage, 2013. The concept of α-outliers in structured data situations, in: C. Becker, R. Fried, S. Kuhnt (Eds.), Robustness and Complex Data Structures. Festschrift in Honour of Ursula Gather. Springer, Berlin, pp. 85–101.

[25] R. Li, M. Chow, 2005. Evaluation of reproducibility for paired functional data. J. Multivariate Anal. 93, 81–101.

[26] M. Lichman, 2013. UCI machine learning repository. URL: http://archive.ics.uci.edu/ml.

[27] S. López-Pintado, J. Romo, 2009. On the concept of depth for functional data. J. Am. Stat. Assoc. 104, 718–734.

[28] S. López-Pintado, Y. Sun, J.K. Lin, M.G. Genton, 2014. Simplicial band depth for multivariate functional data. Adv. Data Anal. Classif. 8, 321–338.

[29] K. Mosler, Y. Polyakova, 2012. General notions of depth for functional data. URL: http://arxiv.org/pdf/1208.1981.

[30] A. Nieto-Reyes, 2011. On the properties of depth for functional data, in: F. Ferraty (Ed.), Recent Advances in Functional Data Analysis. Springer, Berlin, pp. 239–244.

[31] A. Nieto-Reyes, H. Battey, 2014. On the definition of functional depth. URL: http://arxiv.org/pdf/1410.5686.

[32] R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

[33] J.O. Ramsay, B.W. Silverman, 2005. Functional Data Analysis. 2nd ed., Springer, New York.

[34] J.O. Ramsay, H. Wickham, S. Graves, G. Hooker, 2014. fda: Functional data analysis. URL: http://CRAN.R-project.org/package=fda. R package version 2.4.4.

[35] A. Rehage, S. Kuhnt, 2014. An angle-based functional depth measure for outlier detection, in: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna, pp. 233–238.

[36] R.S. Tsay, 2015. MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models. URL: http://CRAN.R-project.org/package=MTS. r package version 0.33.

[37] D. Valencia, R. Lillo, J. Romo, 2014. Spearman coefficient for functions, in: E.G. Bongiorno, E. Salinelli, A. Goia, P. Vieu (Eds.), Contributions in infinite-dimensional statistics and related topics. Società Editrice Esculapio, Bologna, pp. 269–272.

[38] B. Williams, M. Toussaint, A.J. Storkey, 2008. Modelling motion primitives and their timing in biologically executed movements, in: J. Platt, D. Koller, Y. Singer, S. Roweis (Eds.), Advances in Neural Information Processing Systems 20. Curran Associates, Inc., pp. 1609–1616.

[39] Y. Zuo, R. Serfling, 2000. General notions of statistical depth function. Ann. Statist. 28, 461–482.