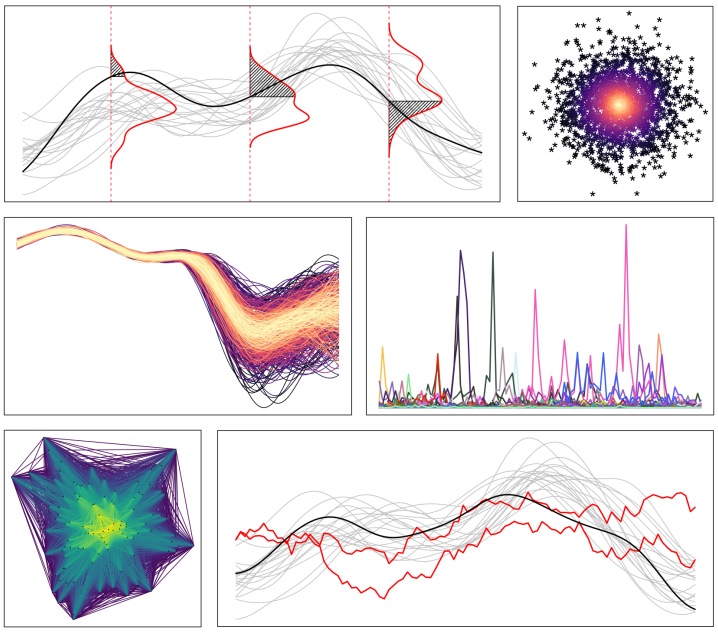
GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Functional Data Depth Overview And An Extreme Value Theory Modification

Master Thesis



Author:
Gordon Schücker

First academic advisor:
Jun.-Prof. Dr. Andrea Krajina

Second academic advisor:
Prof. Dr. Tatyana Krivobokova

Faculty of Mathematics and Computer Science
Institute for Mathematical Stochastics

2017

Statutory Declaration

I assure the single handed composition of this master thesis to be only supported by declared resources. The thesis in this form or in any other form has not been submitted to an examination body and has not been published.

....

Place, Date Signature

Acknowledgements

Without the help of numerous persons, this master thesis would not have been possible and I therefore thank everyone that directly and indirectly contributed to its success.

Foremost, I want to thank Jun.-Prof. Dr. Andrea Krajina for agreeing to supervise this thesis. Her patience and guidance have been invaluable to me. I want to particularly thank her for answering my countless questions and always proving detailed feedback. I am also very thankful to Prof. Dr. Tatyana Krivobokova for her acceptance of her role as second advisor.

I am deeply honored to always have had my familys love and support throughout the years. My parents, Ulrike and Thomas Schücker have always been an inspiration to me and I will forever be thankful to them for enabling me to come thus far.

I am grateful to all my many friends for sharing with me this somewhat longer than expected, but worthwhile journey. The incredible skills I have learned in and along my studies will hopefully help me navigate this interesting and ever changing world. I am eager to find out.

Contents

1 Introduction

In recent years, the substantial decrease in the cost of sensors and computing power has lead to vast data accumulation. Data can now be recorded with such precision that many can be considered as functional data. As a result, nowadays, functional data are common in virtually any field, including life sciences, engineering, physical sciences, finance and social sciences.

To handle this complex data, the statistical field of functional data analysis has seen accelerated advances with the development of many novel tools. One such tool relates to the notion of data depth, a concept that has been introduced to extend univariate approaches based on order to first multivariate data, then also to functional and multivariate functional data. The resulting center-outward ordering has been proven successful at revealing diverse features of the underlying distribution and have led to the development of numerous depth constructions for functional data and useful applications thereof, such as data classification or outlier detection. A visualization of a *data depth for functional data*, we also refer to as *functional depth function*, is presented in Figure 1.1.

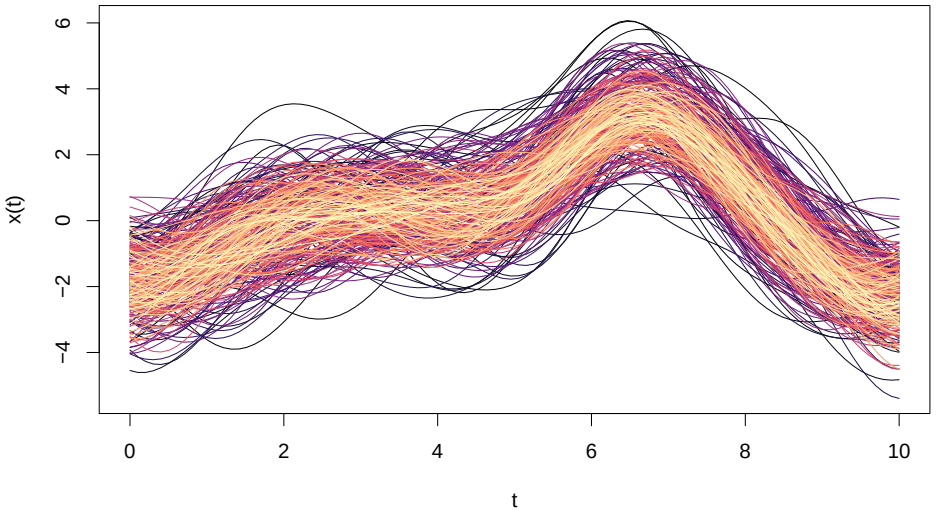


Figure 1.1: Visualization of a functional depth function (the functional random projection depth) applied to a sample of stochastic processes generated by a Gaussian process[1] ($n = 300$). Colors go from dark for extremes to light for deeper curves.

At first, this thesis analyses the numerous depth constructions for functional data that have been

[1]The data has been simulated by applying the Gaussian process regression approach introduced in Ebden (2015). We discuss it and give more details about the general way we simulate the resulting type of Gaussian processes in the Appendix. Additionally, all **R** codes are made available. see also the Appendix.

proposed in the literature. For this, we attempt to address the following three questions asked in Zuo and Serfling (2000):

i) What properties should a well defined statistical depth function possess?
ii) Do existing depth functions possess all desired properties?
iii) What constructive approaches lead to attractive depth functions?

To determine the desirable properties, we use the theoretical properties listed in Claeskens, Hubert, Slaets, and Vakili (2014), Nieto-Reyes and Battey (2014) and Narisetty and Nair (2015) to define a general definition of a depth function for functional data. We then group the presented functional depth functions into three distinct groups and study their performance with respect to the introduced properties. To further differentiate between the depth functions, some additional properties discovered across the literature are also included in the analysis and the number of functional depth functions deemed practical is reduced to only four.

The second focus of this thesis is to consider an extreme value theory based modification for functional depth functions. However, since many functional depth functions rely on a depth function for multivariate data as a building block, we consider these depth functions first. Of special interest is the random Tukey depth which is known to perform poorly near and outside of the convex hull of the data. This entices us, on the one hand, to question the method used in Cuesta-Albertos and Nieto-Reyes (2007) to determine the number of required random projections for the random Tukey depth and subsequently reevaluate this number. On the other hand, it inspires us to modify existing depth functions applying the random Tukey depth with the help of extreme value theory using the approach from Einmahl, Li, and Liu (2015). By analyzing this modified depth function, surprising similarities with the random projection depth introduced in Cuevas, Febrero, and Fraiman (2007) are also found. Based on the insights gained from the multivariate case, we conclude that the only depth function for functional data to possibly benefit from an extreme value theory based modification is the depth function we renamed "the multivariate functional random projection depth", originally introduced in Cuevas et al. (2007).

In summary, this thesis is structured as follows: In Chapter 2 we cover the history of data depth, lay out the notation and present a definition of a depth function for multivariate and functional data. In Chapter 3 we present existing depth constructions for functional data. Subsequently, in Chapter 4 we analyze their theoretical properties together with helpful visualizations. In Chapter 5 we introduce concepts of extreme value theory and apply these to modify the multivariate functional random projection depth. To finish, the performance of the resulting depth function is assessed and a conclusion is given.

2 Data Depth Concept

2.1 History

Contrary to the univariate case, there exists no straightforward method to order a multivariate data set. The first attempt to address this issue was made with the introduction of the Mahalanobis distance (Mahalanobis, 1936). It measures the distance of a point $x \in \mathbb{R}^K$, $K \in \mathbb{N}$, with respect to a distribution P by estimating the number of standard deviations the point x is away from the mean. While this method is well suited for analyzing elliptical distributions, it can fall short in the case of more general multivariate data that are not necessarily related to a parametric distribution. In 1975, Tukey proposed an intuitive way to visualize bivariate data using the method of *convex hull peeling*. The idea consists of starting with a polygon enveloping the most outward data points. These are all assigned a rank of one and discarded. The process is then repeated to assign all points of the next envelop a rank of two, and so on, until all points have been ranked.[1] In the end, the result can be represented as *depth contour levels*, as illustrated in Figure 2.1. Clearly, the underlying distribution plays an important role since the closer a point is to the center of the data set, the higher its assigned rank is. Hence this method provides a non-parametric center-outward ordering and gives rise to the term *depth*. Since the concept of centrality can be interpreted in many ways, a myriad of multivariate depth definitions have been proposed since, e.g. Oja (1983), Liu (1990), Fraiman et al. (1999), Vardi and Zhang (2000) and Zuo (2003). While the steps involved in the calculations can vary greatly from one depth function to another, they have in common that they all achieve a projection of points in $\mathbb{R}^K$ to $\mathbb{R}^+$, $K \in \mathbb{N}$. This dimensionality reduction enables the use of univariate concepts on multivariate data for statistical inference, for example, Jörnsten (2004), Ghosh and Chaudhuri (2005) and Li et al. (2012) proposing depth based data classifications, Donoho and Gasko (1992) and Mosler (2002) outlier detection and finally Liu and Singh (1997) hypothesis testing.

The next natural step is to apply the concept of statistical depth to functional data, the main focus of this thesis. Here the data is typically observed in the form of curves $x_1(t), x_2(t), \ldots, x_n(t)$, $n \in \mathbb{N}$, over time t in some interval $U \subseteq \mathbb{R}$. While specific depth functions for multivariate data can be expanded to the functional case, the resulting depth functions often do not incorporate the topological features of functional data. First, while in practice the functions are usually observed in vector form, the evaluation points can differ in location and amount from one curve to another and contrarily to the multivariate case, they cannot be permuted. Also due to the dependence structure of the stochastic process, one often observes autocorrelated data which complicates the application of some standard

[1] Note that this is the idea behind the *Tukey depth* introduced in Tukey (1974). The Tukey depth itself is, however, defined differently. See (3.2) in the following Chapter 3.

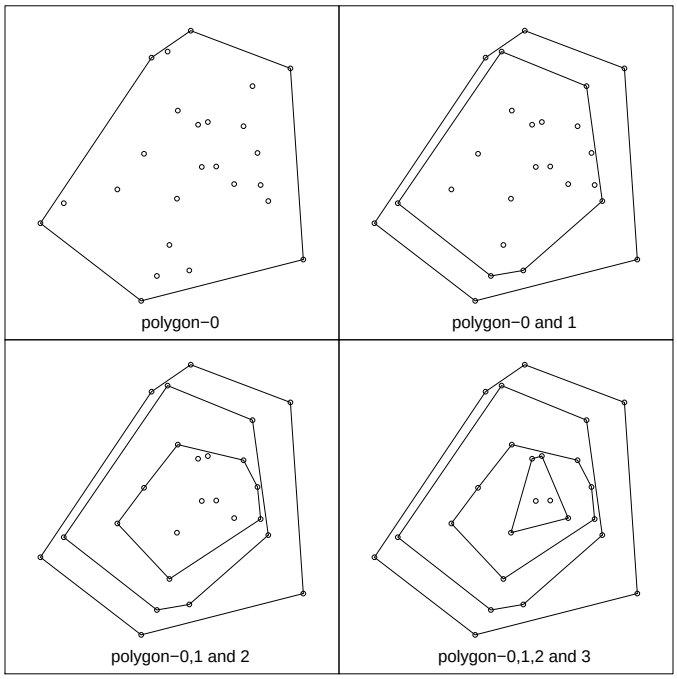


Figure 2.1: Tukey's convex hull peeling visualized for bivariate data.

multivariate procedures. Finally, in a more practical sense, one might also be interested to capture the derivative of the curves into the analysis, which further motivates a new class of data depth functions.

Depth functions for functional data introduced in the literature are, among many, the *integrated depth* (Fraiman and Muniz, 2001), the *functional random projection depth* (Cuevas, Febrero, and Fraiman, 2007), the *half-region depth* and the *modified half-region depth* (López-Pintado and Romo, 2011), the *band depth* and *modified band depth* (López-Pintado and Romo, 2009), the *extremal depth* (Narisetty and Nair, 2015) and the *total variation depth* (Huang and Sun, 2016). As for the case on $\mathbb{R}^K$, the functional data depths have lead to numerous applications, for example, constructing trimmed functional means in Fraiman and Muniz (2001) by using the integrated depth approach, classifying functional data in López-Pintado and Romo (2006) with the help of the band depth or Narisetty and Nair (2015) and Huang and Sun (2016) introducing rules for outlier detections.

Functional depth functions have recently also been extended to multivariate functional data sets with Hlubinka, Gijbels, Omelka, and Nagy (2015) and Claeskens, Hubert, Slaets, and Vakili (2014) proposing a generalized version of the integrated depth. In the latter paper, a special case with the use of the Tukey depth as marginal depth is presented. In López-Pintado, Sun, Lin, and Genton (2014), the *simplicial band depth* and the *modified simplicial band depth* are introduced to generalize in some form, respectively, the band depth and the modified band depth. Finally a bivariate random projection depth, the depth function we later modify in Chapter 5, is developed by Cuevas et al. (2007) where the bivariate functional data consists of univariate functional data (first dimension) together with their derivatives (second dimension).

To better understand functional depth functions, it is helpful to first conceptualize some desired

theoretical properties.

2.2 Theoretical Properties

2.2.1 Multivariate Case

While many depth functions have been developed ad hoc to fulfill a specific data analysis goal, their potential reapplication has sparked an interest to study their theoretical properties. In Liu (1990), it is shown that the simplicial depth possesses useful properties, which are subsequently formalized in Zuo and Serfling (2000). For their reproduction below, we denote by $\mathcal{P}$ a class of distributions on the Borel sets of $\mathbb{R}^K$, $K \in \mathbb{N}$ and by $P = P_{\boldsymbol{X}}$ the distribution of the random variable $\boldsymbol{X}$, although the subscript $\boldsymbol{X}$ is suppressed when not explicitly needed.

Definition 2.1. (Zuo and Serfling, 2000)
The bounded and non-negative mapping $D(\cdot, \cdot) : \mathbb{R}^K \times \mathcal{P} \longmapsto \mathbb{R}$ is called a *statistical depth function* if it satisfies the following properties:

- $M1$. *Affine invariance.* $D(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, P_{\boldsymbol{A}\boldsymbol{X}+\boldsymbol{b}}) = D(\boldsymbol{x}, P_{\boldsymbol{X}})$ holds for any $\boldsymbol{x} \in \mathbb{R}^K$, any non-singular matrix $\boldsymbol{A} \in \mathbb{R}^{K \times K}$ and any $\boldsymbol{b} \in \mathbb{R}^K$ and any $P_{\boldsymbol{X}} \in \mathcal{P}$.
- $M2$. *Maximality at center.* $D(\boldsymbol{\theta}, P) = \sup_{\boldsymbol{x} \in \mathbb{R}^K} D(\boldsymbol{x}, P)$ holds for any $P \in \mathcal{P}$ having a unique center of symmetry $\boldsymbol{\theta} \in \mathbb{R}^K$ with respect to some notion of symmetry.
- $M3$. *Monotonicity relative to the deepest point.* For any $P \in \mathcal{P}$ having deepest point $\boldsymbol{\theta}$, $D(\boldsymbol{x}, P) \leq D(\boldsymbol{\theta} + \alpha(\boldsymbol{x} - \boldsymbol{\theta}), P)$ holds for all $\alpha \in [0, 1]$ and any $\boldsymbol{x} \in \mathbb{R}^K$.
- $M4$. *Vanishing at infinity.* $D(\boldsymbol{x}, P) \to 0$ for any $\boldsymbol{x} \in \mathbb{R}^K$ with $\|\boldsymbol{x}\| \to \infty$, for each $P \in \mathcal{P}$, where $\|\cdot\|$ is the Euclidean norm.

Since this definition is generalized to the functional space in near identical form in the next subsection, we discuss the implications of the above properties there.

2.2.2 Functional Case

For the functional framework, we first lay out the notation in detail. We denote by $\mathcal{C}(U)^K$ the set of K-variate continuous functions on an interval $U \subseteq \mathbb{R}$. Further, we denote by $\mathcal{P}$ class of distributions on the Borel sets of $\mathcal{C}(U)^K$, where the Borel sets are defined in the usual sense, see e.g. Billingsley (1968) and Billingsley (2012). To avoid new symbols, we keep $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(K)})$ to denote in its most general form a K- variate stochastic processes with distribution $P_{\boldsymbol{X}}$ on $\mathcal{P}$. However, if $K = 1$ we do not use bold characters. We further use $\boldsymbol{x} = (x^{(1)}, \ldots, x^{(K)})$ to describe a K-variate functional realization. The vector obtained by evaluating the stochastic process $\boldsymbol{X}$ at a point $t \in U$ is denoted by

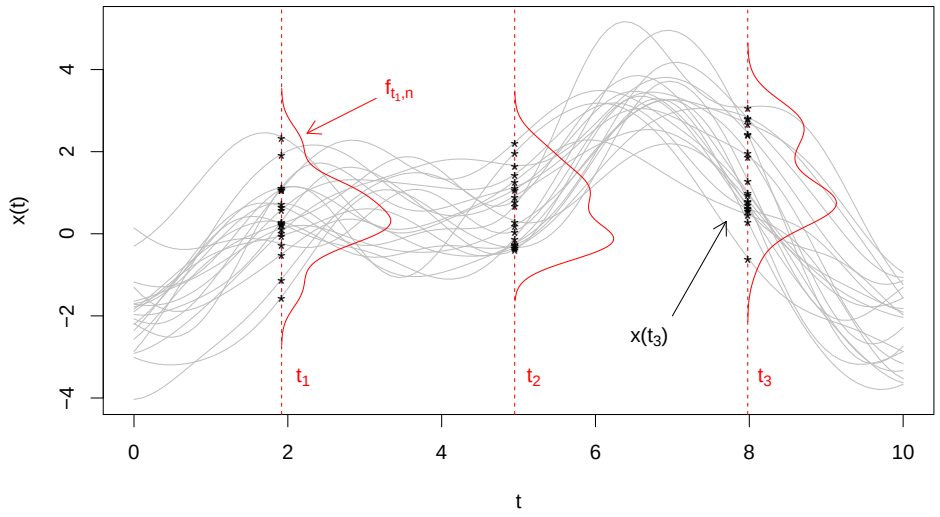


Figure 2.2: Sample of $n = 20$ univariate realizations $\{x_1, \ldots, x_{20}\}$ of independent Gaussian processes. The points in black represent the marginal vectors $x(t_1), x(t_2)$ and $x(t_3)$ with their corresponding kernel density estimates (in red).

$\boldsymbol{X}(t) = (X^{(1)}(t), \ldots, X^{(K)}(t)) \in \mathbb{R}^K$ and we denote its corresponding marginal distribution by $P_{\boldsymbol{X}(t)}$. For the sample $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$ we further denote the empirical distribution by

$$P_{\boldsymbol{X},n}(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{x_i} \in A\}, \tag{2.1}$$

for a Borel set $A \subseteq \mathcal{C}(U)^K$, and the empirical marginal distribution by

$$P_{\boldsymbol{X}(t),n}(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{x_i}(t) \in A\}, \tag{2.2}$$

for Borel set $A \subseteq \mathbb{R}$, $t \in U$, where $\mathbb{1}\{\boldsymbol{x_i} \in A\}$ is the classical indicator function, i.e.

$$\mathbb{1}\{\boldsymbol{x_i} \in A\} := \begin{cases} 1 & \text{if } \boldsymbol{x_i} \in A \\ 0 & \text{if } \boldsymbol{x_i} \notin A. \end{cases}$$

If the subscript $\boldsymbol{X}$ is not needed, we here too suppress it to obtain P, P_t, P_n and $P_{t,n}$, respectively. Furthermore we use F to denote the cumulative distribution function corresponding to P, F_t to denote the marginal cumulative distribution of F at $t \in U$ corresponding to P_t and we denote by F_n and $F_{n,t}$ the empirical counterparts of, respectively, F and F_t. We use f to denote the density function of F and f_t to denote the marginal density function for $t \in U$. Finally, we denote by $f_{t,n}$ the marginal kernel density estimator of f_t. Figure 2.2 further helps to illustrate the used concepts and notation.

We can now define a statistical depth for multivariate functional data as follows.

Definition 2.2. The bounded and non-negative mapping $D(\cdot, \cdot) : \mathcal{C}(U)^K \times \mathcal{P} \longmapsto \mathbb{R}^+$ is called a

multivariate functional depth function if it satisfies the following properties:

P1. *Affine invariance.* $D(\boldsymbol{A}\boldsymbol{x}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}, P_{\boldsymbol{A}\boldsymbol{X}_{(ct+d)}+\tilde{\boldsymbol{X}}_{(ct+d)}}) = D(\boldsymbol{x}, P_{\boldsymbol{X}})$ holds for any constants $c \in \mathbb{R} \setminus \{0\}$, $d \in \mathbb{R}$, any vector of functions $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathcal{C}(U)^K$ with the interval $U := [l, u]$, any $P_{\boldsymbol{X}}, P_{\tilde{\boldsymbol{X}}} \in \mathcal{P}$ and any non-singular matrix $\boldsymbol{A} \in \mathbb{R}^{K\times K}$. Further, $\boldsymbol{A}\boldsymbol{x}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}$ is the function $s \longmapsto \boldsymbol{A}\boldsymbol{x}(\frac{s-d}{c}) + \tilde{\boldsymbol{x}}(\frac{s-d}{c})$ with $s \in S = [cl + d, cu + d]$, $\tilde{\boldsymbol{x}}_{(ct+d)}$ is the function $s \longmapsto \tilde{\boldsymbol{x}}(\frac{s-d}{c})$, $s \in S$, $\boldsymbol{A}\boldsymbol{X}_{(ct+d)} + \tilde{\boldsymbol{X}}_{(ct+d)}$ is the stochastic process $\{\boldsymbol{A}\boldsymbol{X}(\frac{s-d}{c}) + \tilde{\boldsymbol{X}}(\frac{s-d}{c}) : s \in S\}$ and $\tilde{\boldsymbol{X}}_{(ct+d)}$ is the stochastic process $\{\tilde{\boldsymbol{X}}(\frac{s-d}{c}) : s \in S\}$.

P2. *Maximality at center.* $D(\boldsymbol{\Theta}, P_{\boldsymbol{X}}) = \sup_{\boldsymbol{x}\in\mathcal{C}(U)^K} D(\boldsymbol{x}, P_{\boldsymbol{X}})$ holds for any $P_{\boldsymbol{X}} \in \mathcal{P}$ having a unique center of functional symmetry $\boldsymbol{\Theta} \in \mathcal{C}(U)^K$.

P3. *Monotonicity relative to the deepest point.* For any $P_{\boldsymbol{X}} \in \mathcal{P}$ having a deepest point $\boldsymbol{\Theta}$, $D(\boldsymbol{x}, P_{\boldsymbol{X}}) \leq D(\boldsymbol{\Theta} + \alpha(\boldsymbol{x} - \boldsymbol{\Theta}), P_{\boldsymbol{X}})$ holds for any $\boldsymbol{x} \in \mathcal{C}(U)^K$ and for any $\alpha \in [0, 1]$.

P4. *Null at the boundary.* It holds that $\lim_{m\to\infty} D(\boldsymbol{x}_{\boldsymbol{m}}, P_{\boldsymbol{X}}) = 0$ for any $P_{\boldsymbol{X}} \in \mathcal{P}$ and for all sequences of functions $\boldsymbol{x}_{\boldsymbol{m}} \in \mathcal{C}(U)^K$ with $\lim_{m\to\infty} |x_m^{(k)}(t)| = \infty$ for at least one $t \in U$ and one $k \in \{1, \ldots, K\}$.

Property P1 is a generalization from $\mathbb{R}^K$ to the multivariate functional space and states that any data transformation through scaling and translation does not affect the value of the depth. This property is in practice of particular interest when one seeks to recenter the functional data around the null function through subtraction of a mean or median. While often the value of the depth itself is of interest, in some applications this property could be weakened by only requiring the ranking of the data to be invariant under the mentioned transformations.

Property P2 presents some difficulties as the lack of a unique notion of symmetry, even in the multivariate case, prevents any tangible way of confirming this property in practice. For this reason, we use the interpretation of symmetry proposed in Nieto-Reyes and Battey (2014), where Property P2 is satisfied, if the depth function of a standard Gaussian process is maximal at the zero mean function $\Theta \equiv 0$, since it is known that for a Gaussian distribution many symmetry concepts coincide.

Property P3 together with Property P2 ensures that the depth function assigns successively lower or equal depth values to successively larger envelopes around the deepest point Θ, i.e. no local features are allowed to be reflected in the depth.

Property P4 ensures compactness of depth-trimmed regions and allows - by using truncation arguments - to establish almost sure convergence of sample depth to population versions.

Note that our definition of a depth functions differs with the one given in Claeskens et al. (2014) in the definition of Property P4. In that paper, a property called *vanishing at infinity* is given, where the sequence $\boldsymbol{x}_{\boldsymbol{m}}$ requires $\lim_{m\to\infty} |x_m^{(k)}(t)| = \infty$ for *almost all* $t \in U$ and one $k \in \{1, \ldots, K\}$. Thus, this is a significantly weaker property than our definition of Property P4. We illustrate the difference of the two properties in Figure 2.3. Also note the name *null at the boundary* is borrowed from Narisetty and Nair (2015), where a similar definition of Property P4 is given.

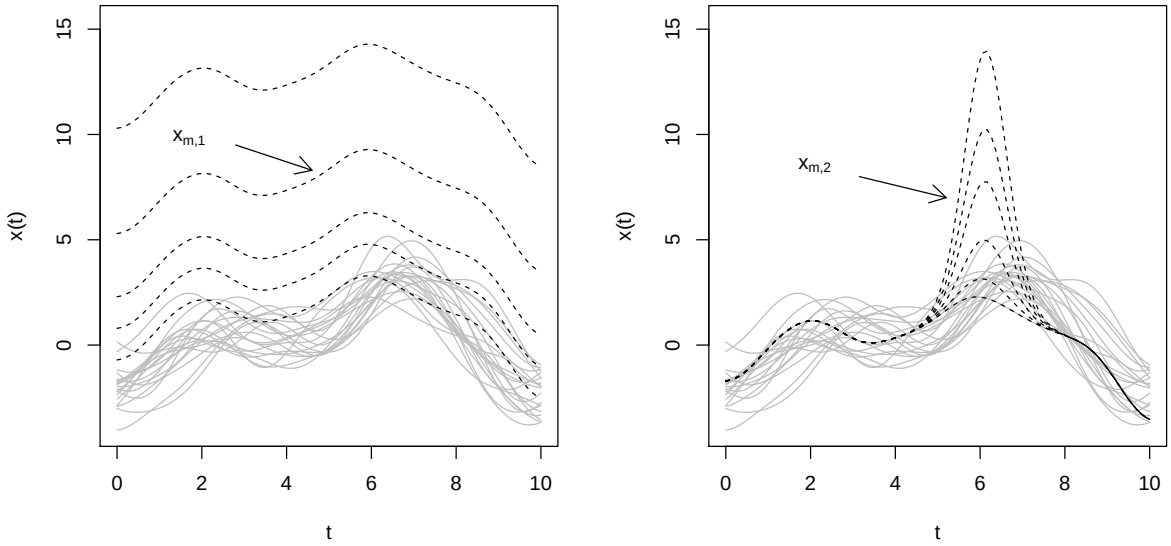


Figure 2.3: Illustration of property P4 versus the *vanishing at infinity* property. Realizations $x_1, \ldots, x_{20}$ of univariate Gaussian processes are plotted together with series of functions $x_{m,1}$ and $x_{m,2}$ (dashed lines). Since $|x_{m,1}(t)| \to \infty$ for all $t \in U$, this would lead, under both properties, to $\lim_{m\to\infty} D(x_{m,1}, P_n) = 0$. However, since $|x_{m,2}(t)| \to \infty$ for one $t \in U$, $\lim_{m\to\infty} D(x_{m,2}, P_n)$ might no be zero under the *vanishing at infinity* property, but would be zero under Property P4.

Additionally, a property called *receptivity to convex hull width across the domain* is proposed in Nieto-Reyes and Battey (2014). For functional data one often observes some subintervals of U where variability is low and where one might not expect to gain much information. Hence, for the computation of the depth function one might not want to assign a lot of weight to the curves on these intervals, as overlapping is usually significant. We therefore informally define that a depth fulfills Property P5, if it naturally incorporates the mentioned characteristic. Figure 2.4 illustrates a data set of $n = 800$ observations with varying amplitudes. We apply the extremal depth (left plot) and the total variation depth (right plot) to it. We notice that the total variation depth seems to rank the functions with focus on the right side of the interval $U = [0, 10]$, where most of the variability is observed, as the deep curves (light curves) are centered only on this interval. The extremal depth seems to be insensitive to this variability as we observe "dark edges" (i.e. extreme curves) on the whole interval $U = [0, 10]$. In particular on the high variability interval, the band of deep curves is quite large. According to our definition, this means that the extremal depth does not satisfy Property P5 while the total variation depth does.

Another important aspect is the *continuity in* P. In general we expect the empirical depth function to converge almost surely to its population counterpart, i.e. that the estimator $D(\cdot, P_n)$ converges P-almost surely to $D(\cdot, P)$. However, functional data is inherently only partially observable as we do not have access to a complete sample due to the limitation of data collection. Instead of observing the curves $\boldsymbol{x_i}$ for $i = 1, \ldots, n$, we only have access to a finite number of points $\big(t_{i,j}, \boldsymbol{x_i}(t_{i,j})\big) \in U \times \mathbb{R}^K$,

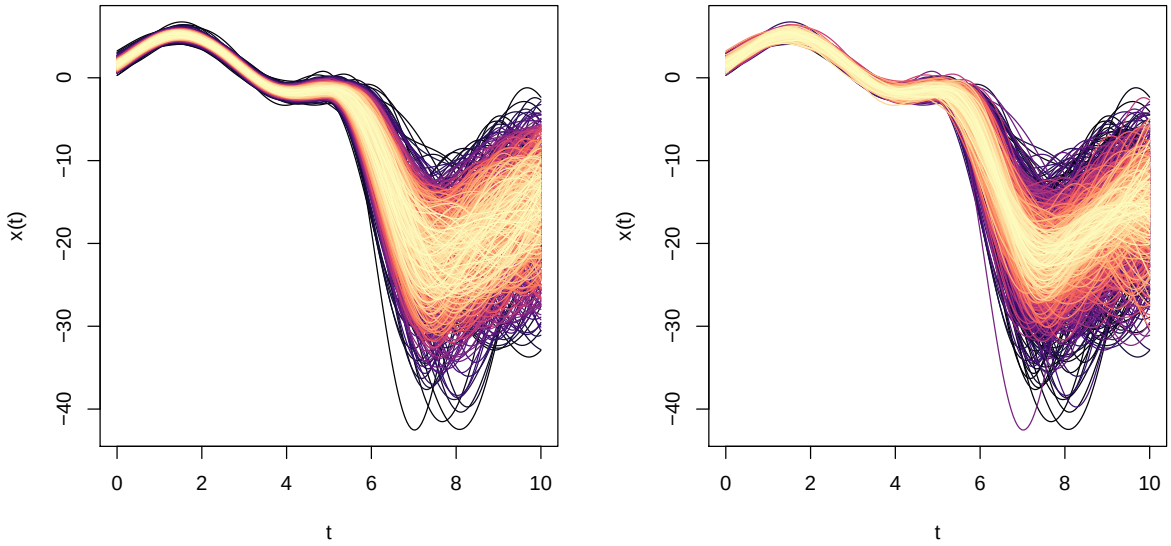


Figure 2.4: Data set of $n = 800$ observations drawn from a modified Gaussian process with varying amplitudes. On the left side the modified half is applied and on the right side the total variation depth is applied. Depths go from dark for extremes to light colors for deeper curves.

for $i = 1, \ldots, n$, $j = 1, \ldots, T_i$, $T_i \in \mathbb{N}$ and $t_{i,j} \in U$. Hence the empirical distribution P_n of the sample $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$ is not available. This issue can be addressed by interpolating the observed function points to obtain the curves $\tilde{\boldsymbol{x}}_{\boldsymbol{i}}$, for $i = 1, \ldots, n$, and thus obtain a corresponding new empirical distribution estimator $\tilde{P}_n$. If the reconstruction leads to $\tilde{P}_n$ converging almost surely to P, then the *continuity in P* property is satisfied. For simplicity, we assume from here on any observation to already be in its reconstructed functional form. Figure 2.5 shows a sample of observations x_1 and x_2 and their functional reconstruction.

Remark 2.3. As stated, Property P2 and Property P3 ensure that no local features are reflected in the definition of data depth. To address this issue, local depths for functional data have been developed, such as the h-depth (Cuevas et al., 2007) or the *local half-region depth* (Agostinelli, 2015). These depths allow the detection of partial centers by inspecting for each curve a neighborhood given by a band with specific width above and below the studied curve (this is comparable to inspecting a neighborhood of a point within in a specific radius in $\mathbb{R}^K$). The choice of the bandwidth is critical for a successful identification of these partial centers. Note that if the bandwidth tends to infinity, the usual (global) depth is often recovered.

One drawback of local depth functions is that the monotonicity of the depth function along any ray originating from the deepest point is lost. This might not allow the straightforward application of some developed methods for statistical inference, e.g. outlier detection. Additionally, if a poor bandwidth is chosen, the quality of the analysis might suffer. On the other hand, if the studied data is known to be multimodal, for example bimodal univariate data, global depths will in most cases missidentify the deepest point and thus lead to false conclusions.

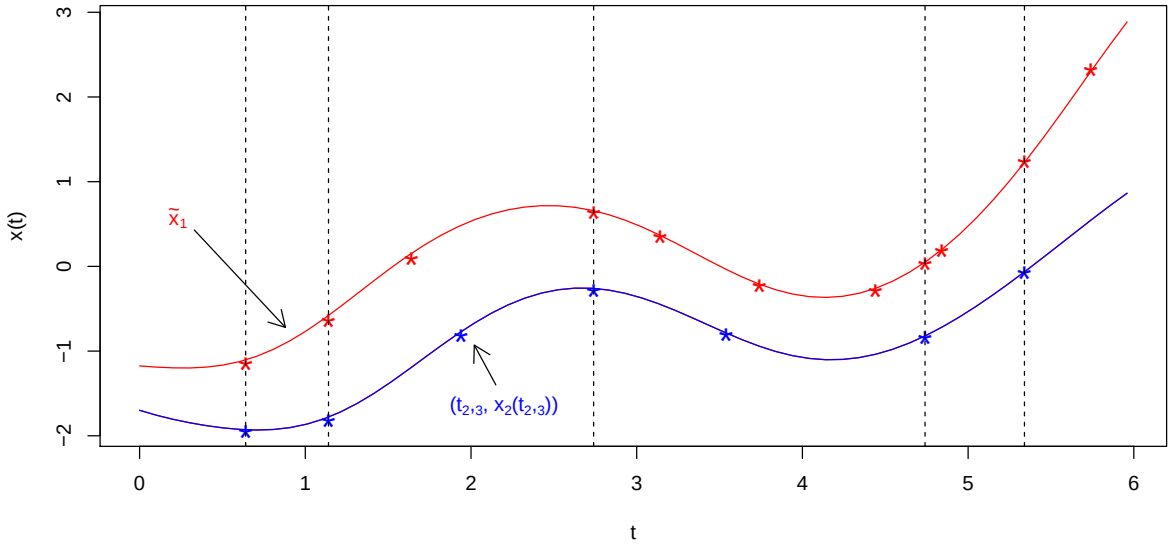


Figure 2.5: Two observations $\{(t_{1,1}, x_1(t_{1,1}), \ldots, (t_{1,11}, x_1(t_{1,11})\}$ (red points) and $\{(t_{2,1}, x_2(t_{2,1}), \ldots, (t_{2,7}, x_2(t_{2,7})\}$ (blue points) with $T_1 = 11$ and $T_2 = 7$ and corresponding functional reconstructions $\tilde{x}_1$ and $\tilde{x}_2$ (curves in respective colors). Note that only a few points of x_1 and x_2 are observed for a same value of $t_{i,j}$, $i = 1, 2$, $j = 1, \ldots, T_i$ (dashed lines).

A further consideration of local depth functions is however beyond the scope of this thesis.

3 State of the Art

We now introduce some selected depth functions in a detailed manner, on the one hand to provide a deeper insight into the current state of research on functional depths and on the other hand to familiarize the reader with data depth definitions that are candidates for possible extreme value theory modifications in Chapter 5. We begin with depth functions for multivariate data.

3.1 Depth Functions for Univariate and Multivariate Data

Assume for the following depths a given probability distribution P and a corresponding cumulative distribution function F on $\mathbb{R}^K$, $K \in \mathbb{N}$. Define the empirical distribution by

$$P_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{x_i} \in A\}, \tag{3.1}$$

for a Borel set $A \subseteq \mathbb{R}^K$ and $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n} \in \mathbb{R}^K$ realizations of independent random variables from distribution P.

3.1.1 Tukey Depth

Definition 3.1. The *univariate Tukey depth* (Tukey, 1974)[1] of $x \in \mathbb{R}$ with respect to P is defined by

$$D_{T1}(x, P) := \min\{P(-\infty, x], P[x, \infty)\} = \min\{F(x), 1 - F(x)\}. \tag{3.2}$$

Each point x is assigned the minimal probability which can be attained in the two closed half-spaces bounded by x, as illustrated in Figure 3.1.

To obtain the empirical version of (3.2), replace the unknown P by its empirical counterpart P_n introduced in (3.1).

Definition 3.2. The empirical multivariate Tukey depth of $x \in \mathbb{R}$ with respect to P_n is defined by

$$D_{T1}(x, P_n) := \min\{P_n(-\infty, x], P_n[x, \infty)\} = \min\{F_n(x), 1 - F_n(x)\}, \tag{3.3}$$

where $F_n(x)$ is the empirical cumulative distribution function of x with respect to the sample $\{x_1, \ldots, x_n\}$

[1] Tukey himself gave it the name *half-space depth*.

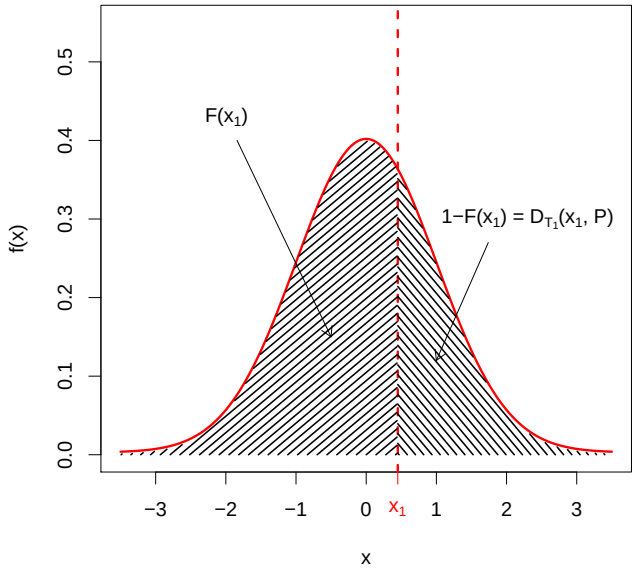


Figure 3.1: Probability of the two half-spaces bounded by x_1 (dashed areas) with the depth of x_1 corresponding to the minimum of the two. The curve in red represents the density function.

of realizations of random variables from distribution P defined by

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\}. \tag{3.4}$$

For the multivariate case, the analogous half-space approach is used.

Definition 3.3. Given $\boldsymbol{v} \in \mathbb{R}^K$, $K > 1$, let $\Pi_{\boldsymbol{v}}$ be the projection of $\mathbb{R}^K$ onto the one dimensional subspace generated by $\boldsymbol{v}$ and let $P_{\boldsymbol{v}} := P \circ \Pi_{\boldsymbol{v}}^{-1}$ be the marginal distribution on this subspace. With this the Tukey depth of $x \in \mathbb{R}^K$ with respect to P is defined by

$$D_T(\boldsymbol{x}, P) := \inf_{\boldsymbol{v}\in\mathbb{R}^K} \{D_{T1}(\Pi_{\boldsymbol{v}}(\boldsymbol{x}), P_{\boldsymbol{v}})\}, \tag{3.5}$$

where D_{T1} is the univariate Tukey depth introduced in (3.2).

The depth is the infimum of the half-space depths of all one-dimensional projections of $\boldsymbol{x}$ with respect to the corresponding marginal distributions $P_{\boldsymbol{v}}$. Two such possible projections are depicted in Figure 3.2.

To apply the Tukey depth in (3.5) to data, we replace the distribution P by its empirical counterpart P_n.

3.1.2 Random Tukey Depth and Random Projection Depth

Although the Tukey depth performs well in theory, it has a high computation time, in particular for higher dimensions. A proposed solution in Cuesta-Albertos and Nieto-Reyes (2007), is to replace the

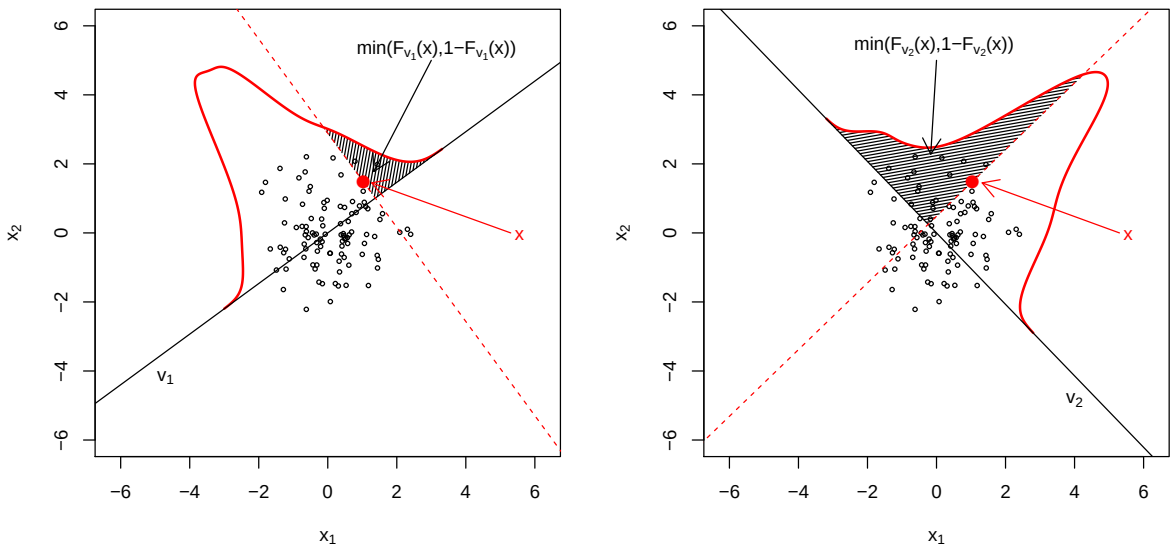


Figure 3.2: Two marginal kernel density functions (curves in red) obtained through projections onto $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ and respective univariate depths calculated for the point $\boldsymbol{x}$ (dashed areas).

infimum in (3.5) by a minimum over a finite number of randomly chosen projections.

Definition 3.4. Let ν be an absolutely continuous distribution on $\mathbb{R}^K$. *The random Tukey* depth of $\boldsymbol{x} \in \mathbb{R}^K$ with respect to P and based on k random vectors chosen with distribution ν is

$$D_{T,k,\nu}(\boldsymbol{x}, P) := \min_{i=1,\dots,k} \{D_{T1}(\Pi_{\boldsymbol{v_i}}(\boldsymbol{x}), P_{\boldsymbol{v_i}})\}, \tag{3.6}$$

where $\boldsymbol{v_1}, \dots, \boldsymbol{v_k}$ are realizations of independent random vectors with distribution ν and $P_{\boldsymbol{v_i}}$ and $\Pi_{\boldsymbol{v_i}}$ are defined as in Definition 3.3.

To apply the random Tukey depth in (3.6) to data, we calculate the depth with respect to P_n instead of the unknown P. An example is shown in Figure 3.3.

Remark 3.5. The distribution ν is typically chosen to be the uniform distribution on the unit sphere in $\mathbb{R}^K$.

Remark 3.6. In Cuevas et al. (2007), it is proposed to replace the minimum function in (3.6) by the mean function over k. This depth function is called the *random projection depth* and we denote it by $D_{RP,k,\nu}$.

Remark 3.7. Although it might seem counterintuitive to measure the depth based on a random quantity, this method can perform well in practice. If the vectors $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots$ are chosen such that $\{\boldsymbol{v_1}, \dots, \boldsymbol{v_k}\} \subset \{\boldsymbol{v_1}, \dots, \boldsymbol{v_{k+1}}\}$, then it follows from the definition, that

$$D_{T,k,\nu}(\boldsymbol{x}, P) \geq D_{T,k+1,\nu}(\boldsymbol{x}, P) \to D_T(\boldsymbol{x}, P), \quad \text{a.s. for } k \to \infty. \tag{3.7}$$

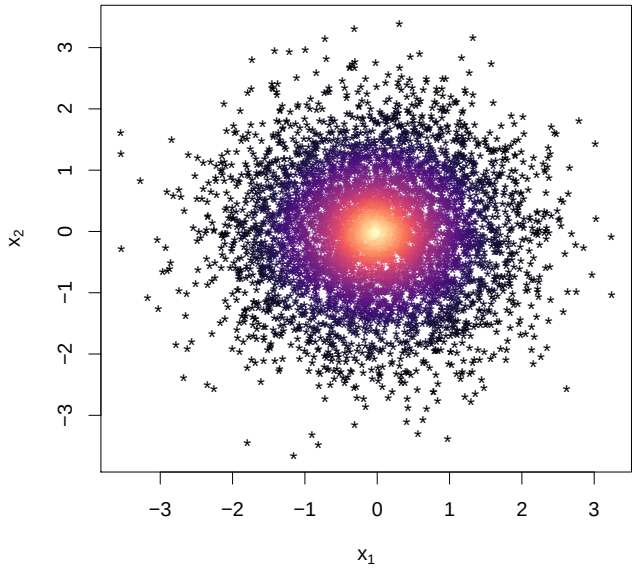


Figure 3.3: Applied random Tukey depth to $n = 5000$ standard bivariate normal realizations, the darker a point is, the more extreme it is estimated to be (low depth value), and vice versa.

Hence, if we choose k large enough, the effect of the randomness in $D_{T,k,\nu}$ becomes negligible. In the investigation carried out in Cuesta-Albertos and Nieto-Reyes (2007), it is concluded that a number of projections of only $k_0 = 10$ would suffice in most common situations. Our findings however show that, often, a much higher value k is needed, particularly in high dimensions. We present our analysis in the following subsection.

Number of Required Projections for the Random Tukey Depth

To determine a good value k_0 of required projections for the random Tukey depth in Cuesta-Albertos and Nieto-Reyes (2007), the performance of $D_{T,k,\nu}$ is analyzed with respect to the *Mahalanobis depth* using simulated data from elliptical distributions, for example, the multivariate normal distribution. The Mahalanobis depth of $\boldsymbol{x} \in \mathbb{R}^K$ is defined by

$$D_{MH}(\boldsymbol{x}, P) := \frac{1}{1 + (\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})},$$

with $\boldsymbol{\mu} \in \mathbb{R}^K$ being the centralization parameter and $\boldsymbol{\Sigma} \in \mathbb{R}^{K \times K}$ the covariance matrix corresponding to P. According to Cuesta-Albertos and Nieto-Reyes (2007), if P is elliptical, $D_T(\cdot, P)$ is close to $D_{MH}(\cdot, P)$. Thus, from (3.7), the larger k is chosen, the higher the resemblance is between $D_{T,k}(\cdot, P)$ and $D_{MH}(\cdot, P)$. This resemblance is measured, for example, by Pearson's correlation coefficient.

Definition 3.8. Let Y_1 and Y_2 be two univariate random variables, then Pearson's correlation coeffi-

cient between Y_1 and Y_2 is defined by

$$\rho(Y_1, Y_2) := \frac{\mathbb{E}[(Y_1 - \mu_{Y_1})(Y_2 - \mu_{Y_2})]}{\sigma_{Y_1}\sigma_{Y_2}} \in [-1, 1], \tag{3.8}$$

with μ_{Y_i} and σ_{Y_i} representing, respectively, the mean and the standard deviation of Y_i, $i = 1, 2$.

Further, let $\boldsymbol{y_1} = (y_{1,1}, y_{1,2}, \ldots, y_{1,n})$ and $\boldsymbol{y_2} = (y_{2,1}, y_{2,2}, \ldots, y_{2,n})$ be two samples in $\mathbb{R}^n$, then Pearson's empirical correlation coefficient between $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ is defined by

$$\hat{\rho}(\boldsymbol{y_1}, \boldsymbol{y_2}) := \frac{\sum_{j=1}^{n}(y_{1,j} - \hat{\mu}_{\boldsymbol{y_1}})(y_{2,j} - \hat{\mu}_{\boldsymbol{y_2}})}{\hat{\sigma}_{\boldsymbol{y_1}}\hat{\sigma}_{\boldsymbol{y_2}}}, \tag{3.9}$$

where $\hat{\mu}_{\boldsymbol{y_i}}$ and $\hat{\sigma}_{\boldsymbol{y_i}}$ represent, respectively, the sample mean and the sample standard deviation of $\boldsymbol{y_i}$, $i = 1, 2$.

With this, the resemblance between $D_{T,k,\nu}(\boldsymbol{X}, P)$ and $D_{MH}(\boldsymbol{X}, P)$ is quantified by

$$r_{k,MH,P} := \rho(D_{T,k,\nu}(\boldsymbol{X}, P), D_{MH}(\boldsymbol{X}, P)),$$

where $\boldsymbol{X}$ is a random vector with distribution P and ρ is Pearson's correlation coefficient defined in (3.8). Note that $D_{MH}(\boldsymbol{X}, P)$ are random variables with realizations in $\mathbb{R}$.

Cuesta-Albertos and Nieto-Reyes (2007) assume that by (3.7), the function $k \longmapsto r_{k,MH,P}$ is increasing[2], which enables us, in theory, to find a value k_0 where improvements become insignificant. But since in practice the distribution P is unknown, the empirical counterpart P_n is used. We hence study the sample correlation coefficient

$$r_{k,MH,P_n} := \hat{\rho}(\boldsymbol{z_{T,k}}, \boldsymbol{z_{MH}}), \tag{3.10}$$

where

$$\begin{aligned} \boldsymbol{z_{T,k}} &:= \big(D_{T,k,\nu}(\boldsymbol{x_1}, P_n), D_{T,k,\nu}(\boldsymbol{x_2}, P_n), \ldots, D_{T,k,\nu}(\boldsymbol{x_n}, P_n)\big) \\ \boldsymbol{z_{MH}} &:= \big(D_{MH}(\boldsymbol{x_1}, P_n), D_{MH}(\boldsymbol{x_2}, P_n), \ldots, D_{MH}(\boldsymbol{x_n}, P_n)\big), \end{aligned}$$

and $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ are realizations of independent random variables $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$ with distribution P. Since $D_{T,k}(\cdot, P)$ and $D_{MH}(\cdot, P)$ are values in $\mathbb{R}$ we have that $\boldsymbol{z_{T,k}}$ and $\boldsymbol{z_{MH}}$ are vectors in $\mathbb{R}^n$ for every $k \in \mathbb{N}$.

However, due to the randomness introduced by P_n, the function $k \longmapsto r_{k,MH,P_n}$ is not necessarily strictly monotone anymore. According to Cuesta-Albertos and Nieto-Reyes (2007) the function would be strictly increasing for small values of k and then at some point stabilize by starting to oscillate. Therefore their estimated k_0 is obtained by

[2] This is not true as we show later, but assume for now that this relationship holds.

$$\hat{k}_0 = \inf_{k \geq 1} \{ r_{k,MH,P_n} > r_{k+1,MH,P_n} \},$$

i.e. $\hat{k}_0$ is the smallest $k \geq 1$ so that the function $k \longmapsto r_{k,MH,P_n}$ decreases for the first time.

However, the issue here is twofold. First, it is missed that $D_{T,k}(\cdot, P_n)$ and $D_{MH}(\cdot, P_n)$ behave differently in the tails. The random Tukey depth assigns a depth of $1/n$ to all points of the sample forming the convex hull. Hence, for large n these assigned depth values are close to zero. Additionally, the larger the dimension K is, the larger the number of points forming the convex hull are.[3] On the other hand, for every point $\boldsymbol{x} \in \mathbb{R}^K$ the Mahalanobis depth assigns a depth value that is proportional to the density function evaluated at $\boldsymbol{x}$, i.e. also for points in the tails. For data points on the convex hull, the corresponding depth values are therefore not necessarily close to zero and as a result $D_{T,k}(\cdot, P_n)$ and $D_{MH}(\cdot, P_n)$ can substantially differ here. This in itself adds natural volatility to the function $k \longmapsto r_{k,MH,P_n}$ and leads to a correlation coefficient that, even in the limit, never reaches 1 and for K large does not even approach 1. Secondly, even if the Mahalanobis depth were a good choice for the comparison, it has not been shown that a beginning in oscillation would immediately imply a lack of further significant improvement. Our findings suggest it does not.

For our analysis, we choose a slightly different approach and examine the correlation between two random Tukey depths. We consider

$$\tilde{r}_{k,k^*,P} := \rho(D_{T,k,\nu}(\boldsymbol{X}, P), D_{T,k^*,\nu}(\boldsymbol{X}, P)),$$

where $D_{T,k,\nu}$ and $D_{T,k^*,\nu}$ represent two random Tukey depths with, respectively, k and k^* random projections drawn independently from one another.

If we fix k^* to a very large value, using the convergence of $D_{T,k,\nu}(\boldsymbol{X}, P)$ to $D_T(\boldsymbol{X}, P)$ described in (3.7), we can approximate $D_T(\boldsymbol{X}, P)$ by $D_{T,k^*,\nu}(\boldsymbol{X}, P)$. It follows that the function $k \longmapsto \tilde{r}_{k,k^*,P}$ approaches one for k approaching k^* from the left, i.e. for $k < k^*$. Due to simulation limitations[4], we choose not to add the restriction that $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots$ are chosen such that $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}\} \subset \{\boldsymbol{v_1}, \ldots, \boldsymbol{v_{k+1}}\}$ as we did in (3.7), but draw $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}\}$ independently of $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_{k+1}}\}$ for computing, respectively, $D_{T,k,\nu}$ and $D_{T,k+1,\nu}$. As a result the function $k \longmapsto \tilde{r}_{k,k^*,P}$ is not necessarily monotone in k.

For application to data we replace P by P_n to obtain

$$\tilde{r}_{k,k^*,P_n} := \hat{\rho}(\boldsymbol{z_{T,k}}, \boldsymbol{z_{T,k^*}}). \tag{3.11}$$

To visualize the difference between the function $k \longmapsto \tilde{r}_{k,k^*,P_n}$ and the function $k \longmapsto r_{k,MH,P_n}$, we plot them together for two simulations with a sample of $n = 100$ random variables from standard multivariate normal distribution with dimensions $K = 4$ and $K = 8$. For $\tilde{r}_{k,k^*,P_n}$ we choose, respectively $k^* = 100\,000$ and $k^* = 1\,000\,000$. We observe that with increasing k, the correlation r_{k,MH,P_n}

[3] For more details about the analysis of the Tukey depth in the tail, we refer to Einmahl et al. (2015) or Dutta et al. (2012).

[4] Unfortunately neither the function "depth" nor the function "depth.space.halfspace" in the **R** package "ddalpha" allow to generate reproducible results, as it is impossible to anchor the realizations $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}\}$ for repeated simulations.

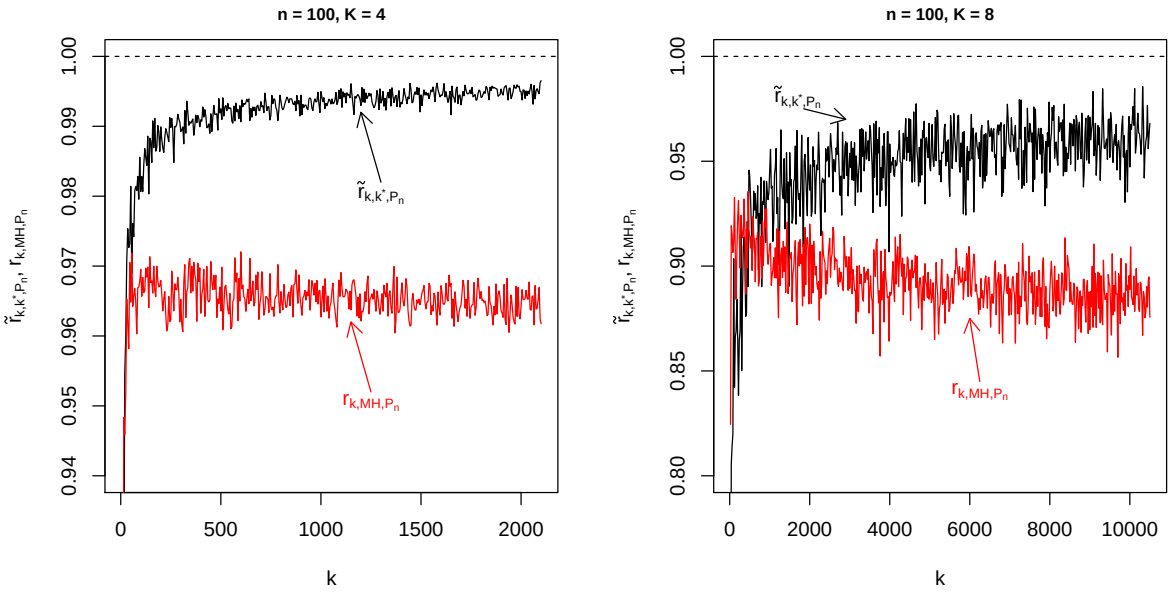


Figure 3.4: The function $k \longmapsto \tilde{r}_{k,k^*,P_n}$ is plotted in black and the function $k \longmapsto r_{k,MH,P_n}$ is plotted in red. In both simulations $n = 100$ random variables from standard multivariate normal distribution were used, on the left with $K = 4$ and on the right with $K = 8$.

tends to decrease, while the correlation $\tilde{r}_{k,k^*,P_n}$ tends to increase, reflecting our previous thoughts. Additionally, we also observe an increase in "volatility" in both functions with increased dimension K.

In order to identify a good value k_0, we carry out a simulation study similar to the one in Cuesta-Albertos and Nieto-Reyes (2007). Here, we consider standard multivariate normal data in dimensions $K = 2, 3, 4$. For $K = 2, 3$ we found out that a value of $k^* = 10\,000$ is large enough and for $K = 4$ we choose $k^* = 100\,000$. We then calculate $\tilde{r}_{k,k^*,P_n}$ for $k = 1, 2, \ldots$ for $m = 1\,000$ Monte Carlo simulations. We denote each Monte Carlo simulation by $\tilde{r}_{k,k^*,P_n,i}$, $i = 1, \ldots, m$. Then, for each $k = 1, 2, \ldots$ we calculate the values

$$\tilde{r}_{k,k^*,P_n,\text{mean}} := \frac{1}{m} \sum_{i=1}^{m} \tilde{r}_{k,k^*,P_n,i},$$

i.e. the average of $\tilde{r}_{k,k^*,P_n,i}$ over all Monte Carlo simulations and

$$\begin{aligned} \tilde{r}_{k,k^*,P_n,\alpha} &:= \tilde{r}_{k,k^*,P_n,(\lfloor \alpha m \rfloor)} \\ \tilde{r}_{k,k^*,P_n,1-\alpha} &:= \tilde{r}_{k,k^*,P_n,(\lceil (1-\alpha) m \rceil)}, \end{aligned}$$

where $\tilde{r}_{k,k^*,P_n,(i)}$ denotes the i-th order statistic of the sample $\{\tilde{r}_{k,k^*,P_n,1}, \ldots, \tilde{r}_{k,k^*,P_n,m}\}$, $\alpha \in [0, 1]$, $\lfloor \cdot \rfloor$ denotes the floor function and $\lceil \cdot \rceil$ denotes the ceiling function. I.e. $\tilde{r}_{k,k^*,P_n,\alpha}$ and $\tilde{r}_{k,k^*,P_n,1-\alpha}$ are,

Dimension		Sample sizes		
		$n = 50$	$n = 100$	$n = 250$
$K = 2$	$k_{0,\alpha}$	6	4	3
	$k_{0,\text{mean}}$	13	9	7
	$k_{0,1-\alpha}$	23	13	10
$K = 3$	$k_{0,\alpha}$	41	22	13
	$k_{0,\text{mean}}$	140	39	22
	$k_{0,1-\alpha}$	421	66	31
$K = 4$	$k_{0,\alpha}$	400	120	50
	$k_{0,\text{mean}}$	3 900	280	75
	$k_{0,1-\alpha}$	22 200	710	106

Table 3.1: Results[5] for different metrics to determine a good value k_0 of required random projections k_0 for the random Tukey depth.

respectively the α and $(1-\alpha)$-percentile of the sample $\{\tilde{r}_{k,k^*,P_n,1}, \ldots, \tilde{r}_{k,k^*,P_n,m}\}$. Further we calculate,

$$k_{0,\text{mean}} = \inf_k \{\tilde{r}_{k,k^*,P_n,\text{mean}} > \rho_0\} \tag{3.12}$$

$$k_{0,\alpha} = \inf_k \{\tilde{r}_{k,k^*,P_n,\alpha} > \rho_0\} \tag{3.13}$$

$$k_{0,1-\alpha} = \inf_k \{\tilde{r}_{k,k^*,P_n,1-\alpha} > \rho_0\} \tag{3.14}$$

where ρ_0 is the correlation level we seek to surpass and is close to one. I.e. $k_{0,\text{mean}}$ is the smallest k for which $\tilde{r}_{k,k^*,P_n,\text{mean}}$ is larger than ρ_0. The values $k_{0,\alpha}$ and $k_{0,1-\alpha}$ are defined similarly. The resulting $k_{0,\text{mean}}$, $k_{0,\alpha}$ and $k_{0,1-\alpha}$ can now be used to asses the function $k \to \tilde{r}_{k,k^*,P_n}$. Finally we carry out the above simulations for $n = 50, 100, 250$ and set $\alpha = 0.05$ and $\rho_0 = 0.99$. The results are summarized in Table 3.1 and Figure 3.5 illustrates two simulations.

From the results, we observe that smaller sample sizes require more projections to attain our set mark of $\rho_0 = 0.99$, in particular for higher dimensions. This is due to the fact that the random Tukey depth performs weakly in the tails and that an increase in the dimension K increases the size of the "tail region". Also, independent of the sample size n, we observe a dramatic increase in required projections, the larger the dimension K gets. The reason behind it is grasped when one considers the uniform distribution on the unit sphere on $\mathbb{R}^K$ we denote by $\mathbb{S}^{K-1}$. Going for example from the bivariate case to the three dimensional case, it is clear than many more possible directions for projections are "possible" in $\mathbb{R}^3$ than in $\mathbb{R}^2$ as in particular $\mathbb{S}^1 \subset \mathbb{S}^2$. In summary, we recommend using a number of required projections of $k_0 = 20$ for $K = 2$, $k_0 = 80$ for $K = 3$ and $k_0 = 800$ for $K = 4$ for $n \geq 100$ and do not recommend the use of the random Tukey depth for low sample sizes. For higher dimensions, further research is required.[6]

[5] In dimension $K = 4$, for k, simulations are carried out in increments of 100 for $n = 50$ and in increments of ten for $n = 100$.

[6] Due to the long running times of our code, a more careful implementation would be required to carry out the presented simulations in higher dimension.

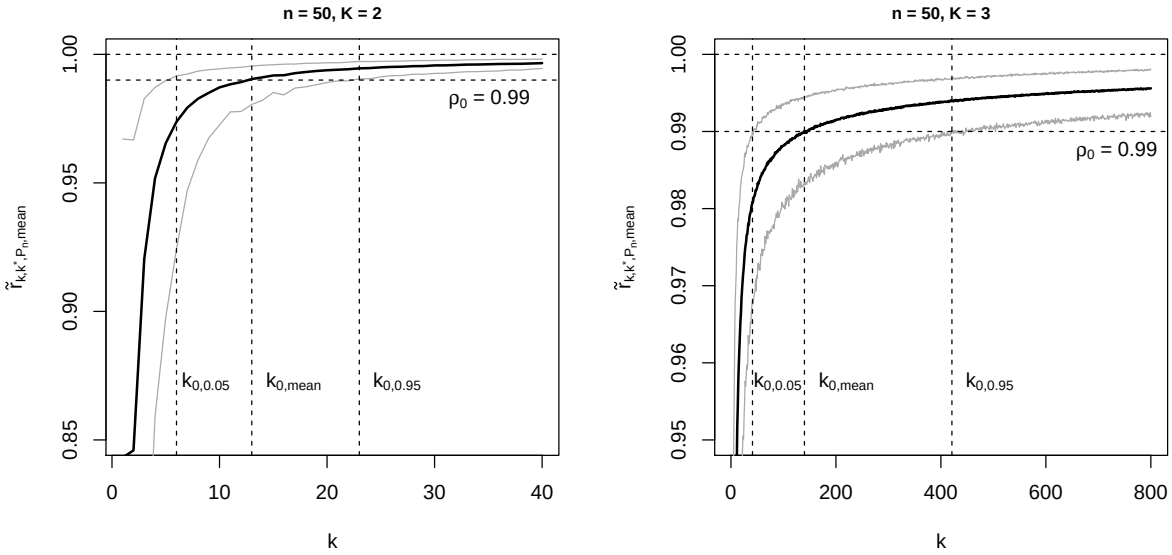


Figure 3.5: The function $\tilde{r}_{k,k^*,P_n,\text{mean}}$ is plotted in black and the functions $\tilde{r}_{k,k^*,P_n,\alpha}$ (lower function) and $\tilde{r}_{k,k^*,P_n,1-\alpha}$ (upper function) are plotted in grey. In both simulations, we choose $n = 50$ and $\alpha = 0.05$. The data is from the standard multivariate normal distribution, with $K = 2$ on the left and $K = 3$ on the right. For the values $k_{0,\text{mean}}$, $k_{0,\alpha}$, $k_{0,1-\alpha}$ vertical dashed lines are drawn and the value ρ_0 is indicated by the horizontal dashed line.

3.1.3 Simplicial Depth

The *simplicial depth* (Liu, 1990) is best introduced in the bivariate case with a sample $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$, where $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n} \in \mathbb{R}^2$ are realizations of independent random variables with distribution P. For any three data points $\boldsymbol{x_i}, \boldsymbol{x_j}$ and $\boldsymbol{x_k}$ from the sample, we can form the closed triangle $\Delta(\boldsymbol{x_i}, \boldsymbol{x_j}, \boldsymbol{x_k})$ with vertices[7] $\boldsymbol{x_i}, \boldsymbol{x_j}$ and $\boldsymbol{x_k}$. Hence from n data points we can generate $\binom{n}{3}$ triangles. For any point $\boldsymbol{x} \in \mathbb{R}^2$ we can then determine in how many data generated triangles it is included to obtain its simplicial depth.

For a better understanding, one can imagine drawing each of the $\binom{n}{3}$ triangles very lightly. The more often a triangle overlaps another one, for example in the center of the data, the darker the lines get. On the other hand, on the periphery, the lines tend to be less concentrated and thus remain light. Figure 3.6 illustrates this process. We now formally introduce the simplicial depth.

Definition 3.9. The simplicial depth of $\boldsymbol{x} \in \mathbb{R}^K$ with respect to P is

$$D_S(\boldsymbol{x}, P) := P\Big\{\boldsymbol{x} \in S\big[\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}\big]\Big\}, \tag{3.15}$$

where $\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}$ are independent random variables with distribution P and $S\big[\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}\big]$

[7] If $\boldsymbol{x_i}, \boldsymbol{x_j}$ and $\boldsymbol{x_k}$ are collinear, then the triangle $\Delta(\boldsymbol{x_i}, \boldsymbol{x_j}, \boldsymbol{x_k})$ is compressed to a line, which is also an accepted case.

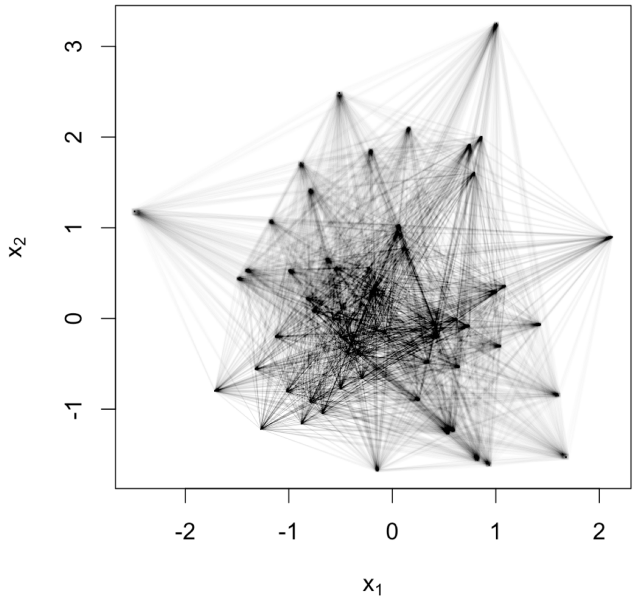


Figure 3.6: Plot of $\binom{n}{3}$ spanned triangles for $n = 60$ simulated standard bivariate Gaussian data points illustrating the simplicial depth. (To make the lines more visible, each triangle was drawn adding a slight jitter.)

denotes the random "simplex"[8] in $\mathbb{R}^K$ with "vertices" $\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}$ defined by

$$S\big[\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}\big] := \Big\{\boldsymbol{Y} : \boldsymbol{Y} = \alpha_1 \boldsymbol{X_1} + \cdots + \alpha_{K+1} \boldsymbol{X_{K+1}} : \sum_{i=1}^{K+1} \alpha_i = 1 : \alpha_i > 0 \; \forall i\Big\}. \tag{3.16}$$

To apply the simplicial depth to data in $\mathbb{R}^K$, we use the $\binom{n}{K+1}$ simplexes generated by the sample $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$ and come to the following definition.

Definition 3.10. The empirical simplicial depth of $x \in \mathbb{R}^K$ with respect to P_n is

$$D_S(\boldsymbol{x}, P_n) := \binom{n}{K+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{K+1} \le n} \mathbb{1}\Big\{\boldsymbol{x} \in S\big[\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}\big]\Big\},$$

where $S\big[\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}\big] \subset \mathbb{R}^K$ is the simplex with vertices $\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}$ defined by

$$S\big[\boldsymbol{x_1}, \ldots, \boldsymbol{x_{K+1}}\big] := \Big\{\boldsymbol{y} \in \mathbb{R}^K : \boldsymbol{y} = \alpha_1 \boldsymbol{x_1} + \cdots + \alpha_{K+1} \boldsymbol{x_{K+1}} : \sum_{i=1}^{K+1} \alpha_i = 1 : \alpha_i > 0 \; \forall i\Big\}. \tag{3.17}$$

It is also interesting to consider the univariate case, $K = 1$, where the simplicial depth in (3.15) reduces to

$$D_{S1}(x, P) := P(x \in \overline{X_1 X_2}), \tag{3.18}$$

[8] Collinearity among $\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}$ is acceptable.

where $X_1 \in \mathbb{R}$ and $X_2 \in \mathbb{R}$ are independent random variables, each with distribution P, and $\overline{X_1 X_2}$ represents the interval $\big[\min\{X_1, X_2\}, \max\{X_1, X_2\}\big]$. For F continuous, we have $P(X_1 \leq X_2) = 1/2$ and can further write

$$\begin{aligned} D_{S1}(x, P) &= \frac{1}{2} P(X_1 \leq x \leq X_2 | X_1 \leq X_2) + \frac{1}{2} P(X_2 \leq x \leq X_1 | X_2 \leq X_1) \\ &= P(X_1 \leq x) P(X_2 \geq x) + P(X_2 \leq x) P(X_1 \geq x) \\ &= 2F(x)[1 - F(x)]. \end{aligned} \tag{3.19}$$

Subsequently, the empirical simplicial depth of $x \in \mathbb{R}$ with respect to P_n is

$$D_{S1}(x, P_n) = 2F_n(x)[1 - F_n(x)],$$

where $F_n(x)$ is the empirical cumulative distribution function of the sample $\{x_1, \ldots, x_n\}$ defined in (3.4).

Remark 3.11. Given the explicit formula of the empirical simplicial depth in (3.10), it is not difficult to see that this depth can be very computationally intensive, especially for larger n and larger K. For more insights, we refer to Afshani, Sheehy, and Stein (2015), where faster running algorithms to compute and approximate the depth are presented, using numerical analysis.

3.2 Depth Functions For Univariate Functional Data

For the following section we study univariate stochastic processes with assumptions and notation from Subsection 2.2.2 for $K = 1$.

Recall that $X_1, \ldots, X_n$ are i.i.d. stochastic processes with distribution P on $\mathcal{C}(U)$, the space of continuous functions on a closed interval $U \subset \mathbb{R}$, P_t is the marginal distribution of P at $t \in U$, P_n is the empirical distribution of the observed functions $x_1, \ldots, x_n$ defined in (2.1) and $P_{t,n}$ is the marginal distribution of P_n at point $t \in U$ defined in (2.2).

Additionally, if an expectation $\mathbb{E}$ is taken we assume it to be with respect to P, unless otherwise stated.

3.2.1 Integrated Depth

Definition 3.12. Let D_M be any data depth for univariate data. The *integrated depth* (Fraiman and Muniz, 2001) of $x \in \mathcal{C}(U)$ with respect to P is defined by

$$D_{ID}(x, P, D_M) := \int_U D_M(x(t), P_t) dt. \tag{3.20}$$

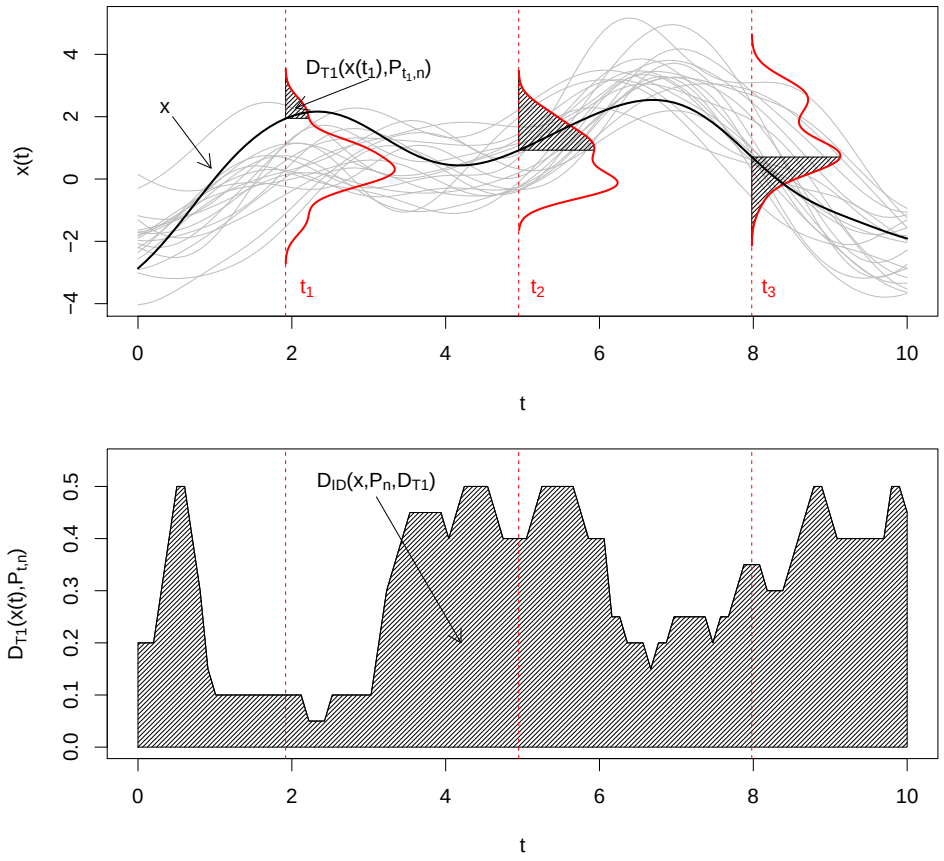


Figure 3.7: Illustration of the integrated depth for $x \in \mathcal{C}([0,10])$ with respect to the sample of $n = 20$ realizations $\{x_1, \ldots, x_{20}\}$ of independent Gaussian processes (in gray). In the top graphic, three marginal kernel density estimates for t_1,t_2 and t_3 are plotted in red with corresponding univariate Tukey depth (dashed areas). In the bottom graphic, the marginal univariate Tukey depth is plotted over U (black line) with the dashed area representing its integral over U, i.e. the final depth value of x.

For all $t \in U$, the marginal depth D_M of $x(t)$ is calculated with respect to the marginal distribution P_t and the resulting depth values are integrated over U to obtain the final value for the integrated depth of the function x. The approach is illustrated in Figure 3.7, with D_M chosen to be the univariate Tukey depth defined in (3.2).

To obtain the empirical version of (3.20) replace P and P_t by, respectively, P_n and $P_{t,n}$.

3.2.2 Functional Random Projection Depth

The *functional random projection depth* (Cuevas, Febrero, and Fraiman, 2007)[9] is similar to the random Tukey depth for multivariate data defined in (3.6) and extended to functional data. Here we use "functional projection angles" to reduce functional data to univariate points on which a depth

[9] In this paper, this depth function is called the *random projection depth* for both the version for multivariate data and the one for functional data. We choose to rename the latter to avoid any confusion.

of choice for univariate data is applied. For the functional random projection depth we expect the trajectories of the stochastic processes $X_1, \dots, X_n$ to be elements of $\mathcal{F} := \mathcal{C}(U) \cap \mathcal{L}^2(U)$, the space of continuous functions intersected with the space of square integrable functions on U.

Definition 3.13. Let D_M be any depth for univariate data and let η be a non-degenerate probability distribution on $\mathcal{F}$. The functional random projection depth of $x \in \mathcal{F}$ with respect to P based on k stochastic processes with distribution η is

$$D_{FRPD,k,\eta}(x, P, D_M) := \frac{1}{k} \sum_{i=1}^{k} D_M(\langle v_i, x \rangle, P_{v_i}), \tag{3.21}$$

where $v_1, \dots, v_k$ are realizations of independent stochastic processes with distribution η and

$$\langle v_i, x \rangle := \int_U v_i(t) x(t) dt \tag{3.22}$$

is the $\mathcal{L}^2$-inner product between v_i and x, $i = 1, \dots, k$. Additionally, P_{v_i} denotes the marginal distribution of P on $\{\langle v_i, y \rangle : y \in \mathcal{F}\}$, $i = 1, \dots, k$.

Remark 3.14. The distribution η is typically chosen to be a non-degenerate stationary Gaussian process on $\mathcal{F}$.

To get the sample analogue of the functional random projection depth in (3.21) we replace P by P_n.

We now present an example and consider the same data as in Figure 3.7. In addition to the $n = 20$ realizations $x_1, \dots, x_{20}$, we simulate two "random projection angle functions" v_1 and v_2 from a stationary Gaussian process. The results are plotted in Figure 3.8. Further, we choose the marginal depth D_M to be equal to the Tukey depth D_{T1} introduced in (3.2).
We define

$$y_{i,j} := \langle v_i, x_j \rangle = \int_U v_i(t) x_j(t) dt$$

and

$$y_{v_i} := \langle v_i, x \rangle = \int_U v_i(t) x(t) dt$$

for $i = 1, 2$, $j = 1, \dots, 20$. Then $P_{v_1,n}$ and $P_{v_2,n}$ are the distributions of the points $y_{1,j}$ and $y_{2,j}$, respectively , $j = 1, \dots, 20$. The final depth is then

$$\begin{aligned} D_{FRPD,2,\eta}(x, P_n, D_{T1}) &= \frac{1}{2}\Big[D_{T1}(y_{v_1}, P_{n,v_1}) + D_{T1}(y_{v_2}, P_{n,v_2})\Big] \\ &= \frac{1}{2}\Big[\min\left\{F_{n,v_1}(y_{v_1}), 1 - F_{n,v_1}(y_{v_1})\right\} \\ &\qquad + \min\left\{F_{n,v_2}(y_{v_2}), 1 - F_{n,v_2}(y_{v_2})\right\}\Big], \end{aligned}$$

the average of the two dashed areas in Figure 3.9.

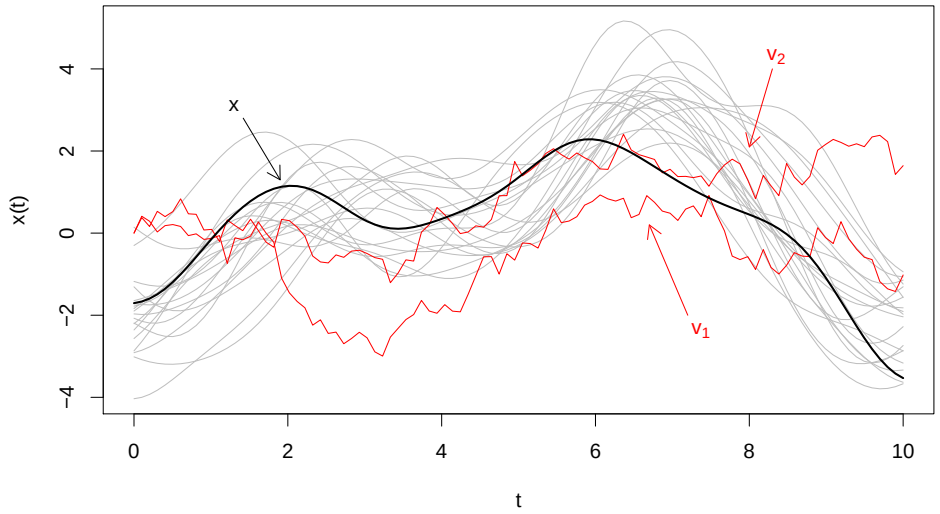


Figure 3.8: Illustration of $n = 20$ realizations of Gaussian processes (in grey), a function x and two "random projection angle functions" v_1 and v_2.

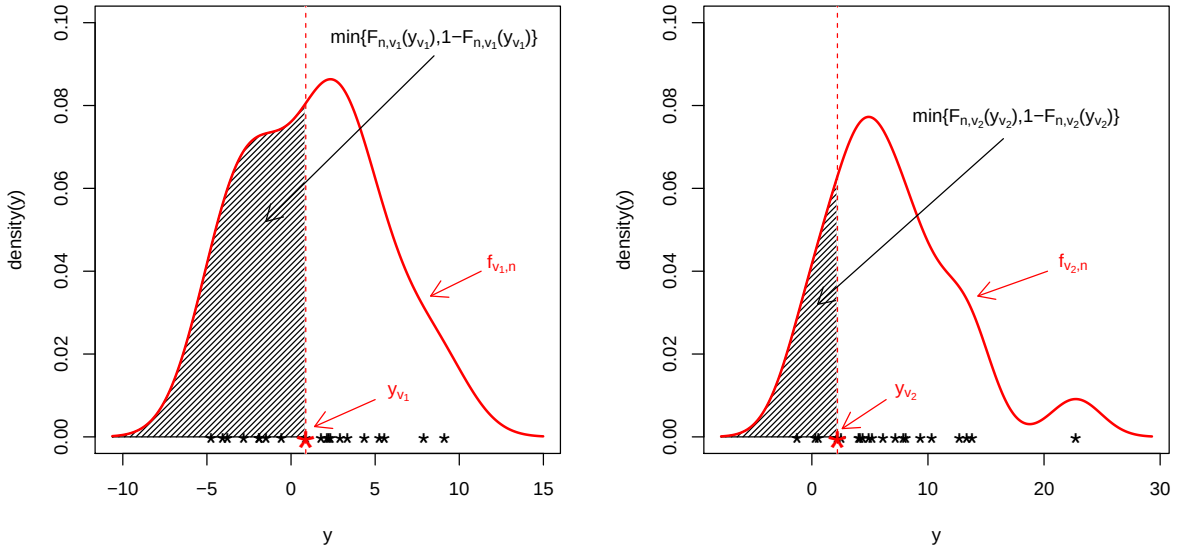


Figure 3.9: Kernel densities (in red) over the points $y_{1,j}$ and $y_{2,j}$, respectively, $j = 1, \ldots, 20$ and corresponding univariate random Tukey depth for y_{v_1} and y_{v_2} (dashed areas).

Remark 3.15. The functional random projection depth can also be modified by replacing the mean in (3.21) by the minimum over $i = 1, \dots, k$. For later examinations, we refer to the two depth functions as $D_{FRPD,k,\eta,\text{mean}}(x, P, D_M)$ and $D_{FRPD,k,\eta,\min}(x, P, D_M)$ if the distinction is required.

Remark 3.16. If $D_M = D_{T1}$ is chosen, as in our previous example, together with the modification of the mean into the minimum as described in the previous remark, then the resulting depth is called the "functional random Tukey depth" in Cuesta-Albertos and Nieto-Reyes (2007) and is, as the name suggests, the generalization to functional data of the random Tukey depth for multivariate data. Note that the functional random Tukey depth "depth.RT" in the **R** package "fda.usc" is in this sense incorrectly implemented as the mean is taken instead of the minimum.

Number of Required Projections for the Functional Random Projection Depth

To determine a good value k_0 of required projections for the functional random projection depth defined in (3.21), we apply the process analogous to the one used in Section 3.1.2. In particular, we adapt (3.11) in the following way.

$$\tilde{r}_{k,k^*,P_n} := \hat{\rho}(\boldsymbol{z_{FRPD,k}}, \boldsymbol{z_{FRPD,k^*}}) \tag{3.23}$$

where

$$\boldsymbol{z_{FRPD,k}} := \Big(D_{FRPD,k,\eta}(x_1, P_n), D_{FRPD,k,\eta}(x_2, P_n), \dots, D_{FRPD,k,\eta}(x_n, P_n)\Big),$$

$$\boldsymbol{z_{FRPD,k^*}} := \Big(D_{FRPD,k^*,\eta}(x_1, P_n), D_{FRPD,k^*,\eta}(x_2, P_n), \dots, D_{FRPD,k^*,\eta}(x_n, P_n)\Big),$$

and $\{x_1, \dots, x_n\}$ is a sample drawn independently from distribution P.

We carry out $m = 1\,000$ Monte Carlo simulations for $n = 50, 100, 250$ and set $\alpha = 0.05$ and $\rho_0 = 0.99$ to calculate $k_{0,\text{mean}}$, $k_{0,\alpha}$ and $k_{0,1-\alpha}$ defined in (3.12) to (3.14) for $\tilde{r}_{k,k^*,P_n}$ defined in (3.23). Analogous to choosing K in the random Tukey depth setting for Table 3.1, we here choose the number of equidistant discretization points $T = 50, 100, 200$ for each function x_i, $i = 1, \dots, n$, i.e. each function x_i is evaluated at equidistant points $t_1, \dots, t_T$. We summarize the results in Table 3.2.

The results seem to indicate that, contrary to the random Tukey depth for multivariate data, neither the number of discretization points T, nor the sample size n seem to much influence the number of required random projections. Hence, it is safe to assume a conservative value $k_0 = 1\,000$ to be a good choice.

3.2.3 Half-Region Depth

Definition 3.17. The *half-region depth* (López-Pintado and Romo, 2011) of $x \in \mathcal{C}(U)$ with respect to P is

$$D_{HRD}(x, P) := \min\{P(X \in H_x), P(X \in E_x)\}, \tag{3.24}$$

Dimension		Sample sizes		
		$n = 50$	$n = 100$	$n = 200$
$T = 25$	$k_{0,\alpha}$	500	500	500
	$k_{0,\text{mean}}$	750	500	500
	$k_{0,1-\alpha}$	1 000	1 000	750
$T = 50$	$k_{0,\alpha}$	500	500	500
	$k_{0,\text{mean}}$	750	500	500
	$k_{0,1-\alpha}$	1 000	750	750
$T = 100$	$k_{0,\alpha}$	500	500	500
	$k_{0,\text{mean}}$	750	500	500
	$k_{0,1-\alpha}$	1 000	750	750

Table 3.2: A good value k_0 of required random projections for the functional random projection depth. Simulations carried out in k increments of 250.

where X is a stochastic process with distribution P. The *hypograph* of x is defined by

$$H_x := \{y \in \mathcal{C}(U) : y(t) \leq x(t), \forall t \in U\}$$

and the *epigraph* of x is defined by

$$E_x := \{y \in \mathcal{C}(U) : y(t) \geq x(t), \forall t \in U\}.$$

For the empirical counterpart we have the following definition.

Definition 3.18. The empirical half-region depth of $x \in \mathcal{C}(U)$ with respect to P_n is

$$D_{HRD}(x, P_n) := \min\left\{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i \in H_x\}, \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i \in E_x\}\right\},$$

where $x_1, \ldots, x_n$ are realizations of independent stochastic processes with distribution P.

The depth is equal to the minimum between the proportion of observations in the hypograph of x and the the proportion of observations in the epigraph of x.

Figure 3.10 illustrates the half-region depth together with the hypograph and epigraph of a curve x. Here $D_{HRD}(x, P_n) = \min\{1/4, 2/4\} = 1/4$.

3.2.4 Modified Half-Region Depth

If the functions in a sample cross one another often, the half-region depth - due to its construction - leads to small depth values. For this reason (López-Pintado and Romo, 2011) also propose a modification of the half-region depth.

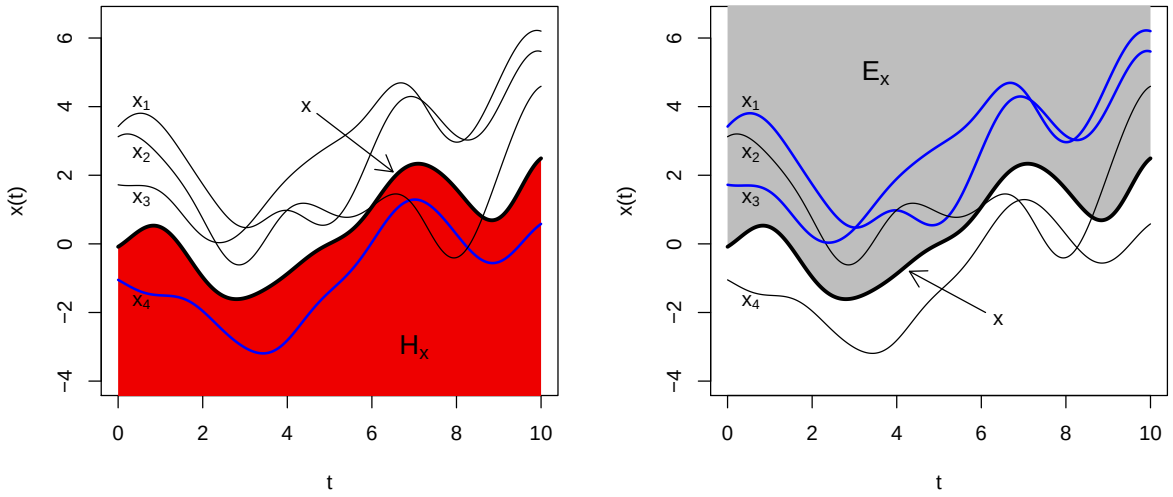


Figure 3.10: Function x and its corresponding hypograph (area in red) and epigraph (area in grey) plotted together with $n = 4$ realizations of stochastic processes, colored in blue if within one of the two latter areas, colored in black if not.

Definition 3.19. The *modified half-region depth* of $x \in \mathcal{C}(U)$ with respect to P is defined by

$$D_{MHRD}(x, P) = \min\left\{\frac{\mathbb{E}\big[\lambda\{t \in U : X(t) \leq x(t)\}\big]}{\lambda(U)}, \frac{\mathbb{E}\big[\lambda\{t \in U : X(t) \geq x(t)\}\big]}{\lambda(U)}\right\}, \tag{3.25}$$

where X is a random process with distribution P and λ is the Lebesgue measure.

In addition to assessing the probability that the stochastic process X is completely in the hypograph or in the epigraph of x, the modified half-region depth also takes into account the probability that the proportion of the stochastic processes X is in either one of the two half-regions.

For the empirical counterpart of (3.25) we have the following definition.

Definition 3.20. The empirical modified half-region depth of $x \in \mathcal{C}(U)$ with respect to P_n is defined by

$$D_{MHRD}(x, P_n) = \min\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\lambda\{t \in U : x_i(t) \leq x(t)\}}{\lambda(U)}, \frac{1}{n}\sum_{i=1}^{n}\frac{\lambda\{u \in U : x_i(t) \geq x(t)\}}{\lambda(U)}\right\},$$

where λ is the Lebesgue measure and $x_1, \ldots, x_n$ are realizations of independent stochastic processes with distribution P.

We reproduce the previous example in Figure 3.11. For 26% of $t \in U$, the observation x_3 is in H_x and for 74% of $t \in U$, the observation x_3 is in E_x. Hence $\lambda\{t \in U : x_3(t) \geq x(t)\}/\lambda(U) = 0.74$ and $\lambda\{t \in U : x_3(t) \leq x(t)\}/\lambda(U) = 0.26$. Thus $D_{MHRD}(x, P_n) = \min\{\frac{1.26}{4}, \frac{2.74}{4}\} = 0.315$.

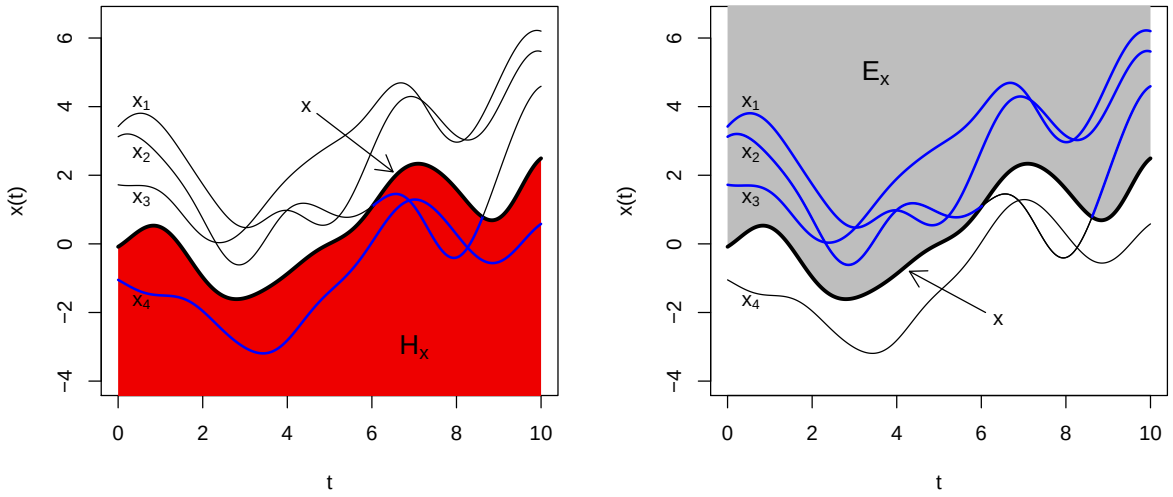


Figure 3.11: Same plot as in Figure 3.10 with the exception that sections of observation x_3 have been colored in blue if these are in the hypograph or in the epigraph of x.

3.2.5 Band Depth

The *band depth* (López-Pintado and Romo, 2009) is similar to the simplicial depth for univariate data introduced in (3.15). For $j \geq 2$, $j \in \mathbb{N}$ and stochastic processes $X_1, \ldots, X_j$ with trajectories in $\mathcal{C}(U)$, we define the "random" j-band by

$$B_j[X_1, \ldots, X_j] := \big\{Y : Y(t) = \alpha_1 x_1(t) + \cdots + \alpha_j x_j(t),\ \forall t \in U : \sum_{k=1}^{j} \alpha_k = 1 : \alpha_k > 0\ \forall k\big\}, \tag{3.26}$$

i.e. it is the set of all possible stochastic processes contained in the convex hull of the stochastic processes $X_1, \ldots, X_j$. We can now define the band depth as following.

Definition 3.21. For a parameter $J \geq 2$ in $\mathbb{N}$, the band depth of $x \in \mathcal{C}(U)$ with respect to P is

$$D_{BD_J}(x, P) = \sum_{j=2}^{J} P\big(x \in B_j[X_1, \ldots, X_j]\big), \tag{3.27}$$

where $X_1 \ldots X_j$ are stochastic processes with distribution P and $B_j[X_1, \ldots, X_j]$ is the random j-band defined in (3.26).

To obtain the empirical counterpart of (3.27) we now, for $j \geq 2$, $j \in \mathbb{N}$ and curves $x_1, \ldots, x_j$ in

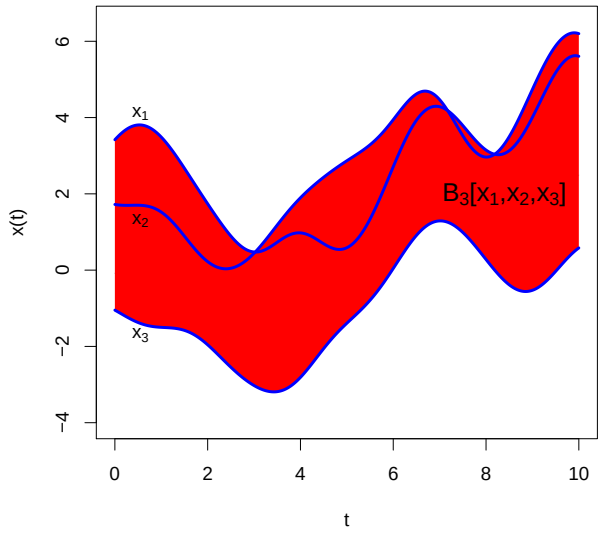


Figure 3.12: Band $B_3[x_1, x_2, x_3]$ drawn from the sample $\{x_1, x_2, x_3\}$ and $j = 3$.

$\mathcal{C}(U)$, define the j-band by

$$B_j[x_1, \dots, x_j] := \big\{ y \in \mathcal{C}(U) : y(t) = \alpha_1 x_1(t) + \cdots + \alpha_j x_j(t), \\ \forall t \in U : \sum_{k=1}^{j} \alpha_k = 1 : \alpha_k > 0 \; \forall k \big\}, \quad (3.28)$$

i.e. it is the set of all curves in $\mathcal{C}(U)$ contained in the convex hull of the curves $x_1, \dots, x_j$. In Figure 3.12 we present an example.

Since for a sample of size n we can form $\binom{n}{j}$ distinct bands, we obtain the following definition of the empirical band depth.

Definition 3.22. For a parameter $J \geq 2$ in $\mathbb{N}$, the empirical band depth of $x \in \mathcal{C}(U)$ with respect to P_n is

$$D_{BD_J}(x, P_n) := \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} \mathbb{1}\big\{ x \in B_j[x_{i_1}, \dots, x_{i_j}] \big\},$$

where $x_1, \dots, x_n$ are independent realizations of stochastic processes with distribution P and $B_j[x_{i_1}, \dots, x_{i_j}]$ is the j-band defined in (3.28) with respect to the curves $x_{i_1}, \dots, x_{i_j}$.

To illustrate the band depth we reconsider the example shown in Figure 3.12. If we set $J = 3$, we obtain four bands $B_2[x_1, x_2]$, $B_2[x_1, x_3]$, $B_2[x_2, x_3]$ and $B_3[x_1, x_2, x_3]$. For the chosen x, the band

depth is then

$$D_{BD_3}(x, P_n) = \frac{1}{4}\Big[\mathbb{1}\{x \in B_2[x_1, x_2]\} + \mathbb{1}\{x \in B_2[x_1, x_3]\} + \mathbb{1}\{x \in B_2[x_2, x_3]\} + \mathbb{1}\{x \in B_3[x_1, x_2, x_3]\}\Big] = 3/4.$$

Remark 3.23. The value for J recommended in López-Pintado and Romo (2009) is $J = 3$ for several reasons: On the one hand, for $J > 3$, calculations can become very computationally intensive. On the other hand, $J = 2$ can lead to some bands having degenerated intervals for some $t \in U$, e.g. $B_2[x_1, x_2]$ in Figure 3.13. These bands would then often not cover any function. However, due to the binomial coefficient in (3.27), even for $J = 3$, the computation time can be very long for large n.

3.2.6 Modified Band Depth

In a one-to-one analogous process to the modification step applied to the half-region depth, we define the *modified band depth* (López-Pintado and Romo, 2009).

Definition 3.24. For a parameter $J \geq 2$ in $\mathbb{N}$, the modified band depth of $x \in \mathcal{C}(U)$ with respect to P is

$$D_{MBD_J}(x, P) = \sum_{j=2}^{J} \mathbb{E}\big[\lambda\{t \in U : x \in B_j[X_1, \ldots, X_j](t)\}\big]/\lambda(U), \tag{3.29}$$

where $X_1, \ldots, X_j$ are independent stochastic processes with distribution P and $B_j[X_1, \ldots, X_j](t)$ is the band defined in (3.26) evaluated at t.

For the empirical counterpart of (3.29) we get the following definition.

Definition 3.25. For a parameter $J \geq 2$ in $\mathbb{N}$, the empirical modified band depth of $x \in \mathcal{C}(U)$ with respect to P_n is

$$D_{MBD_J}(x, P_n) := \sum_{j=2}^{J} \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \cdots < i_j \leq n} \lambda\{t \in U : x(t) \in B_j[x_{i_1}, \ldots, x_{i_j}](t)\}/\lambda(U),$$

where $x_1, \ldots x_n$ are independent realizations of a stochastic processes with distribution P and $B_j[x_{i_1}, \ldots, x_{i_j}](t)$ is the band defined in (3.28) evaluated at t, i.e. an interval.

Remark 3.26. Since the modified band depth uses proportions, the degeneration issue mentioned in Remark 3.23 is not as acute here. Hence the value J recommended[10] in López-Pintado and Romo (2009) is $J = 2$.

Remark 3.27. It is shown in Claeskens et al. (2014), that the modified band depth with $J = 2$ is equivalent to the integrated depth defined in (3.20) with marginal depth equal to the univariate simplicial depth defined in (3.19).

[10] Actually, the **R** function "MBD" provided by (López-Pintado and Romo, 2009) has only been implemented for $J = 2$.

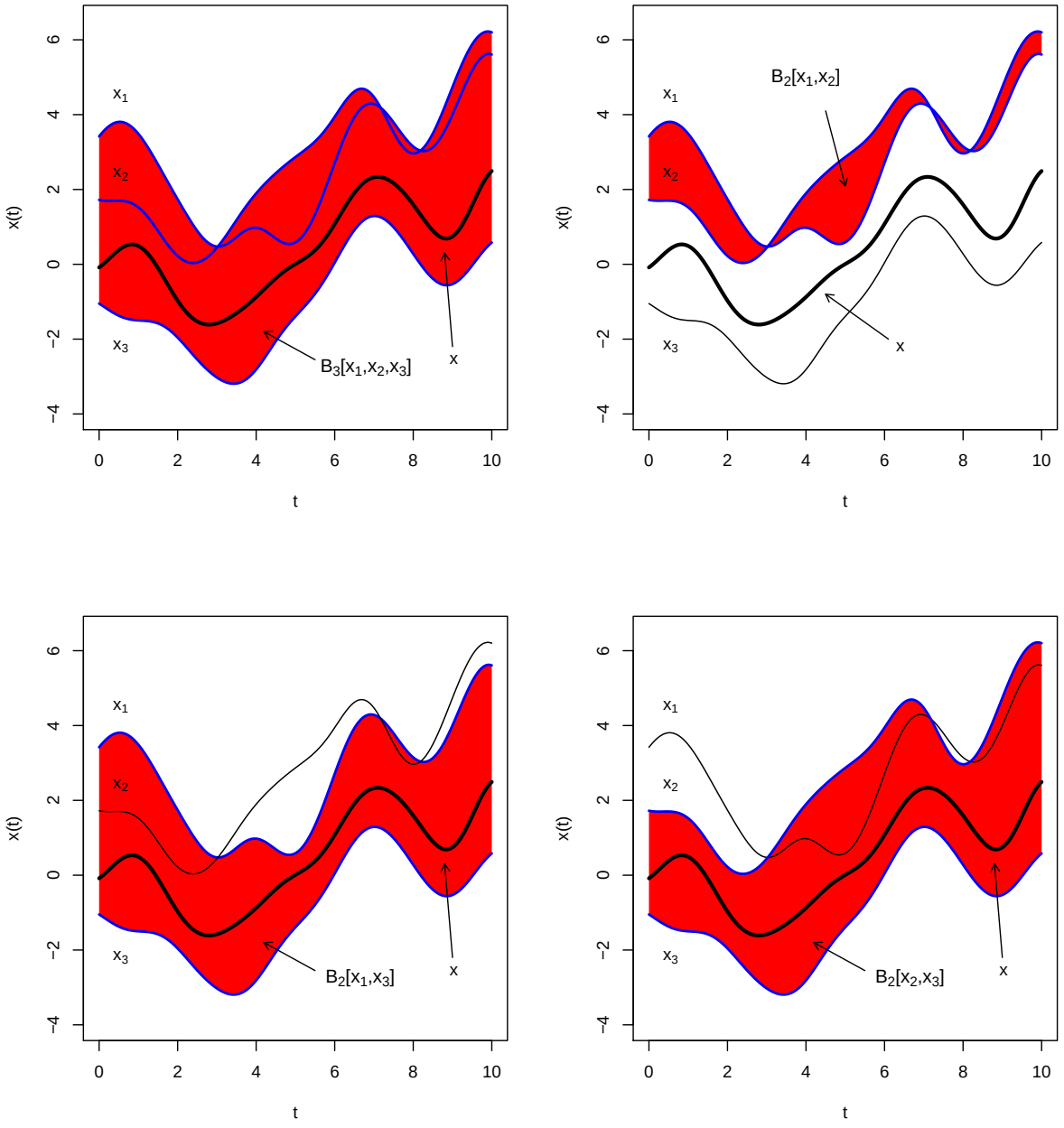


Figure 3.13: All possible bands $B_2[x_1, x_2]$, $B_2[x_1, x_3]$, $B_2[x_2, x_3]$ and $B_3[x_1, x_2, x_3]$ (colored in red) obtained with the sample $\{x_1, x_2, x_3\}$ for $J = 3$ and some specific $x \in \mathcal{C}([0, 10])$. The considered functions for a specific band are colored in blue.

3.2.7 Extremal Depth

To introduce the *extremal depth* (Narisetty and Nair, 2015), as for the integrated depth defined in (3.20), we require a marginal depth D_M for univariate data for $x(t) \in \mathbb{R}$ with respect to the marginal distribution P_t, $t \in U$. Here $D_M = 2D_{T1}$, where D_{T1} is the univariate random Tukey depth defined in (3.2). We now define the function

$$\Phi_x(q, P, D_M) := \int_U \mathbb{1}\{D_M(x(t), P_t) \leq q\}dt, \tag{3.30}$$

for $q \in [0, 1]$. For small q, only "extreme" depth values play a role, hence the name of the depth. To enhance readability going forward, we write

$$\Phi_x(q) := \Phi_x(q, P, D_M).$$

In the next step we use $\Phi_x(q)$ to order a pair of functions $x, y \in \mathcal{C}(U)$. Let

$$q^* := \inf\{q \in [0, 1] : \Phi_x(q) \neq \Phi_y(q)\}. \tag{3.31}$$

If there exists a $\delta > 0$ such that $\Phi_y(q) > \Phi_x(q)$ for all $q \in (q^*, q^* + \delta)$, we say that y is more extreme than x and write $y \prec x$. Note that for $q^* < 1$, such a δ exists as long as $\Phi_x(q)$ and $\Phi_y(q)$ have finitely many crossings. If $q^* = 1$, we say that x is as extreme as y and write $x \sim y$. We can now define the extremal depth as following.

Definition 3.28. The extremal depth of $x \in \mathcal{C}(U)$ with respect to P is defined by

$$D_{ED}(x, P, D_M) := 1 - P(x \prec X),$$

where X is a stochastic process with distribution P.

For the empirical version of the depth, we replace P and P_t by, respectively, P_n and $P_{t,n}$ and obtain the following empirical version of (3.30)

$$\Phi_x(q, P_n, D_M) := \int_U \mathbb{1}\{D_M(x(t), P_{t,n}) \leq q\}dt. \tag{3.32}$$

This time however, $q \in \{0, 1/n, 2/n, \ldots, (n-1)/n, 1\}$, since the range of q is determined by the values taken by the empirical cumulative distribution F_n within the empirical Tukey depth D_{T1} introduced in (3.3). To simplify the notation, we write going forward

$$\Phi_{x,n}(q) := \Phi_x(q, P_n, D_M).$$

With this and the analogous definition of q^* in (3.31), we write the step by step algorithm for ordering the sample $\{x_1, \ldots, x_n\}$ according to the extremal depth.

Extremal Depth Ordering Algorithm

Input: Sample $\{x_1, \dots, x_n\}$, where $x_i \in \mathcal{C}(U)$ for $i = 1, \dots, n$.
Ouput: Ordering of the x_i from most extreme to deepest.

1. For $i = 1, \dots, n$, calculate $D_M(x_i(t), P_{t,n})$ for all $t \in U$.
2. Define $q := (q_1, q_2, \dots, q_{n+1}) := (0, 1/n, 2/n, \dots, (n-1)/n, 1)$.
3. For each $j = 1, \dots, n+1$ and x_i, $i = 1, \dots, n$, calculate $\Phi_{x_i,n}(q_j)$ as in (3.32).
4. Create a start list $L_S := \{x_1, \dots, x_n\}$ and finish list $L_F := \emptyset$.
5. Set $j = 1$.
6.
   6.1. Find the number $k \in \mathbb{N}$ of all $\Phi_{x_i,n}(q_j)$ for which $\Phi_{x_i,n}(q_j) > 0$ holds, with $x_i \in L_S$. Denote $i = i_1, \dots, i_k$ the indices so that

$$\Phi_{x_{i_1},n}(q_j) \geq \Phi_{x_{i_2},n}(q_j) \geq \cdots \geq \Phi_{x_{i_k},n}(q_j) \tag{3.33}$$

   holds with $i_l \in \mathbb{N}$, $l = 1, \dots, k$.

   6.2. If strict inequalities hold for all inequalities in (3.33), go to Step 7.

   6.3. Else, find a distinct pair $i_{l_1}, i_{l_2} \in \{i_1, \dots, i_k\}$ where $\Phi_{x_{i_{l_1}},n}(q_j) = \Phi_{x_{i_{l_2}},n}(q_j)$. Then if $j \neq n+1$, examine $\Phi_{x_{i_{l_1}},n}(q_{j+m})$ and $\Phi_{x_{i_{l_2}},n}(r_{j+m})$, for each $m = 1, 2 \dots$ until the tie is broken or until $j + m = n + 1$.

   If $\Phi_{x_{i_{l_1}},n}(q_{j+m}) > \Phi_{x_{i_{l_2}},n}(q_{j+m})$ for some $m \in \mathbb{N}$, then $x_{i_{l_1}} \prec x_{i_{l_2}}$ and vice versa. If not then $x_{i_{l_1}} \sim x_{i_{l_2}}$. Do this for all pairs for which equality holds until all cases are resolved.
7. Add (in the right order[11]) $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ to L_F and remove them from L_S.
8. If $L_S = \emptyset$, stop and return L_F.
9. Else $j \leftarrow j + 1$ and go to Step 6.

To obtain the depth of a function $x \in \mathcal{C}(U)$, apply the extremal depth ordering algorithm to $\{x_1, \dots, x_n, x\}$.

Definition 3.29. The extremal depth of x with respect to P_n is defined by

$$D_{ED}(x, P_n) := \frac{\#\{i : x \succeq x_i\}}{n},$$

where $x_1, \dots, x_n$ are independent realizations of a stochastic processes with distribution P and $x \succeq x_i$ means that x_i is more or as extreme as x, that is, either $x \succ x_i$ or $x \sim x_i$ holds.

[11] Step 6 provides the right order, but we would need to introduce another index for $x_{i_1}, x_{i_2}, \dots, x_{i_k}$ to write them down in the correct order. We have skipped doing so to keep the notation readable. In R, Step 6 is best implemented using the function "do.call" combined with the function "order", both included in the base package.

	$q_1 = 0$	$q_2 = 1/8$	$q_4 = 3/8$	$q_6 = 5/8$	$q_8 = 7/8$
$\Phi_{x_1,n}(q)$	0.00	**0.50**	**0.50**	**0.67**	1.00
$\Phi_{x_2,n}(q)$	0.00	0.00	**1.00**	1.00	1.00
$\Phi_{x_3,n}(q)$	0.00	0.00	0.00	**0.83**	1.00
$\Phi_{x_4,n}(q)$	0.00	**0.50**	**0.50**	**0.50**	1.00
$\Phi_{x_5,n}(q)$	0.00	**0.33**	0.50	0.67	1.00
$\Phi_{x_6,n}(q)$	0.00	0.00	0.00	**0.33**	1.00
$\Phi_{x_7,n}(q)$	0.00	0.00	**0.50**	1.00	1.00
$\Phi_{x_8,n}(q)$	0.00	**0.67**	1.00	1.00	1.00

Table 3.3: Values of the functions $\{\Phi_{x_1,n}, \Phi_{x_2,n}, \ldots, \Phi_{x_8,n}\}$ for some specific values q.

Remark 3.30. Note that the extremal depth ordering algorithm is limited to ordering of the sample $\{x_1, \ldots, x_n\}$ from the most extreme function (x_i added first to L_F) to the deepest function (x_i added last) and therefore the depth values of x_i only correspond to normalized ranks.

To better understand the algorithm, we provide an example similar to the one in Narisetty and Nair (2015). Consider the sample $\{x_1, x_2, \ldots, x_8\}$ presented in Figure 3.14 together with $\{\Phi_{x_1,n}, \Phi_{x_2,n} \ldots, \Phi_{x_8,n}\}$ for which the corresponding numbers can be found in Table 3.3.

We start with list $L_S := \{x_1, \ldots, x_8\}$ and list $L_F = \emptyset$. We have $q = \{0, 1/8, 2/8, \ldots, 1\}$ and calculate $\{\Phi_{x_1,n}, \Phi_{x_2,n}, \ldots, \Phi_{x_8,n}\}$ for each q_i, $i = 1, \ldots, 8$. We now want to order x_i, $i = 1, \ldots, 8$ and take $j = 1$, i.e. $q_j = 0$, for which $\Phi_{x_i,n}(q_j) = 0$, for all $i = 1, \ldots, 8$. Thus we jump to $q_2 = 1/8$ for which $\Phi_{x_i,n}(1/8) > 0$ for $i = 1, 4, 5, 8$. Here, the highest value of $\Phi_{x_i,n}(1/8)$ is 0.67 and is obtained for $i = 8$, so x_8 is the most extreme function. Then we already have the special case where $\Phi_{x_1,n}(1/8) = \Phi_{x_4,n}(1/8) = 0.5$ (c.f. underlined numbers in Table 3.3). This means we must compare $\Phi_{x_1,n}(q_j)$ and $\Phi_{x_4,n}(q_j)$ for $q_j > 1/8$. Equality still holds for $q_4 = 3/8$, but not anymore for $q_6 = 5/8$. Since $\Phi_{x_1,n}(5/8) = 0.67 > 0.5 = \Phi_{x_4,n}(5/8)$, we end up with $x_1 \prec x_4$. Finally $\Phi_{x_5,n}(1/8) = 0.33$ is the last $\Phi_{x_i,n}(1/8) > 0$ for $i = 1, 4, 5, 8$. Thus we have $x_8 \prec x_1 \prec x_4 \prec x_5$ and add and remove them, respectively, to L_F and from L_S. This yields $L_S = \{x_2, x_3, x_6, x_7\}$.

In the next step we obtain $x_2 \prec x_7$ and then finally $x_3 \prec x_6$. The final ordering then results in $x_8 \prec x_1 \prec x_4 \prec x_5 \prec x_2 \prec x_7 \prec x_3 \prec x_6$ i.e. x_8 is the most extreme observation and x_6 the most central one. This means that the depth of x_8 is equal to $1/8$ and the depth of x_6 is equal to 1.

We finish with an example with a larger sample in Figure 3.15.

Remark 3.31. In Narisetty and Nair (2015), the marginal depth D_M used is chosen as the following

$$D_M\big(x(t), P_t\big) = 1 - \big|P_t\big(X(t) > x(t)\big) - P_t\big(X(t) < x(t)\big)\big|,$$

however, this is just a different formulation of $D_M = 2D_{T1}$, where D_{T1} is the univariate Tukey depth.

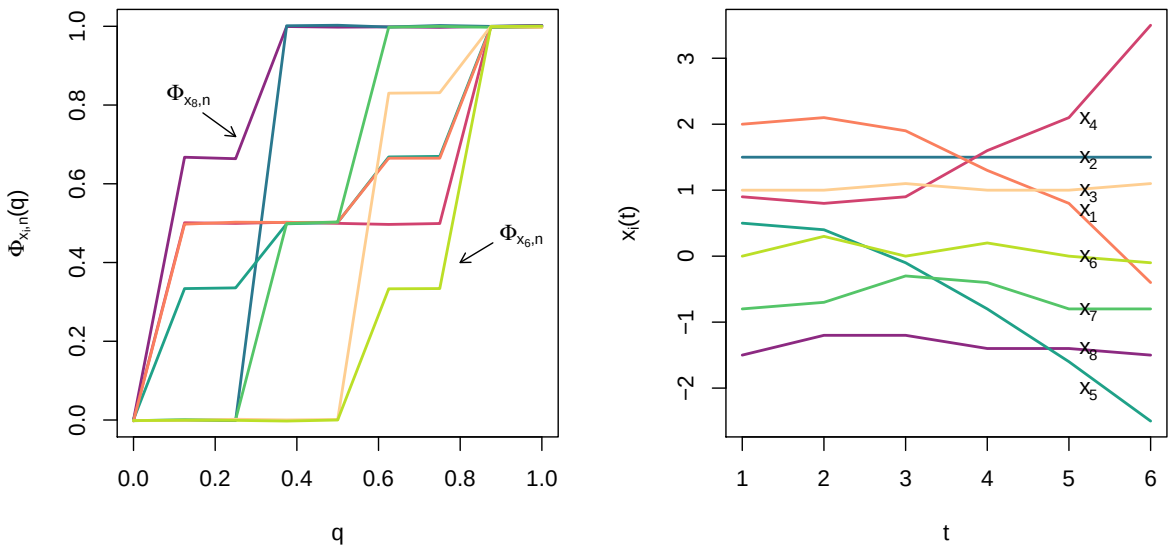


Figure 3.14: On the left, we plot the functions $\Phi_{x_1,n}, \Phi_{x_2,n}, \ldots, \Phi_{x_8,n}$ used for the sorting of the sample $\{x_1, x_2, \ldots, x_8\}$ in the extremal depth algorithm. (A slight jitter is added for better visibility.) Light colors represent deep curves, dark colors extreme curves. On the right, the sample $\{x_1, x_2, \ldots, x_8\}$ is plotted with corresponding colors. The most extreme function is x_8 and the deepest function is x_6.

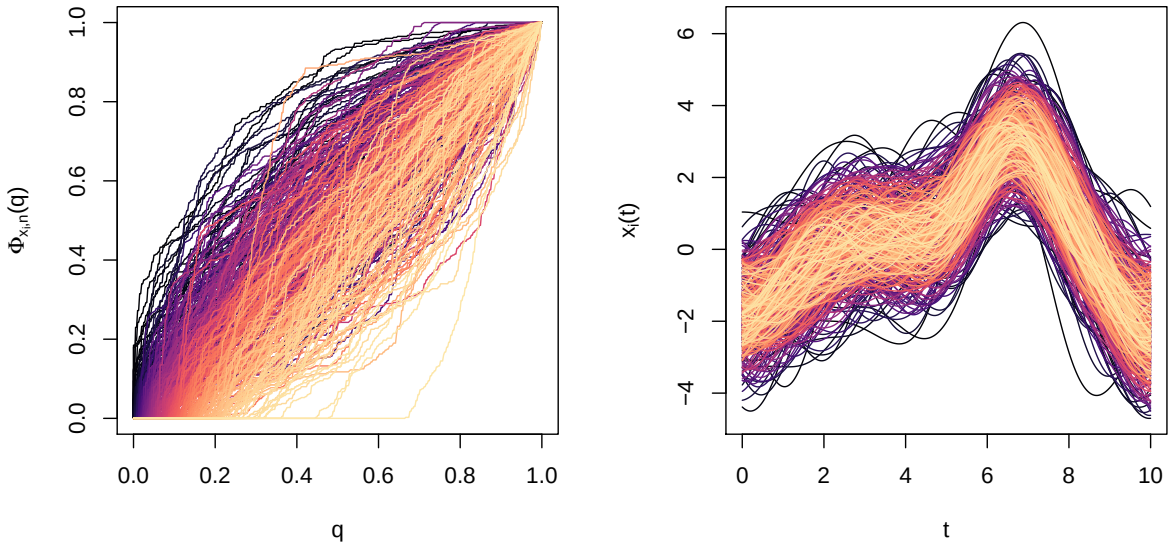


Figure 3.15: Sample $\{x_1, \ldots, x_n\}$, $n = 400$, generated by a Gaussian process. On the left, the functions $q \longmapsto \Phi_{x_i,n}(q)$ are plotted for each corresponding x_i, $i = 1, \ldots, 400$ and on the right, the sample is plotted. Colors are determined the rankings of the $\Phi_{x_i,n}$, $i = 1, \ldots, n$. Light colors represent deep functions and dark colors represent extremes.

3.2.8 Total Variation Depth

The *total variation depth* (Huang and Sun, 2016) is defined as the integrated depth in (3.20) with marginal depth equal to the univariate simplicial depth defined in (3.19) and with an additional weight function. However, in order to later obtain a practical decomposition, some additional notation is used. For a stochastic process X and a function $x \in \mathcal{C}(U)$ define for a given $t \in U$

$$R_x(t) := \mathbb{1}\{X(t) \leq x(t)\}$$

and

$$p_x(t) := \mathbb{E}[R_x(t)] = P_t\big(X(t) \leq x(t)\big) = F_t(x(t)).$$

We then write

$$D_{S1}\big(x(t), P_t\big) = \mathrm{Var}\big(R_x(t)\big) = p_x(t)\big(1 - p_x(t)\big) = F_t(x(t))\{1 - F_t(x(t))\}, \tag{3.34}$$

since $p_x(t)$ is a Bernoulli $\big(F_t(x(t))\big)$ distributed random variable.

We now introduce the total variation depth in the following definition.

Definition 3.32. The total variation depth of $x \in \mathcal{C}(U)$ with respect to P is defined by

$$D_{TVD}(x, P, D_{S1}) := \int_U D_{S1}\big(x(t), P_t\big) w(t) dt, \tag{3.35}$$

where $w(\cdot)\colon U \longmapsto \mathbb{R}^+$ is a non-negative weight function and D_{S1} is the univariate simplicial depth defined in (3.19).

Possible choices for the weight function care, for example

$$w_1(t) :\equiv 1/\lambda(U),$$

where λ is the Lebesgue measure, or

$$w_2(t) := \mathrm{sd}\{X(t)\} / \int_U \mathrm{sd}\{X(s)\} ds,$$

where $\mathrm{sd}\{X(t)\}$ is the the standard deviation of the marginal $X(t)$, assuming it exists for all $t \in U$.

For $s, t \in U$ with $s \leq t$ this notation enables the following decomposition

$$D_{S1}\big(x(t), P_t\big) = \mathrm{Var}\big(R_x(t)\big) = \mathrm{Var}\big(\mathbb{E}\big[R_x(t)|R_x(s)\big]\big) + \mathbb{E}\big[\mathrm{Var}\big(R_x(t)|R_x(s)\big)\big].$$

This means that the total variance of $R_x(t)$ can be decomposed into $\mathrm{Var}\big(\mathbb{E}\big[R_x(t)|R_x(s)\big]\big)$, the variability explained by $R_x(s)$, and $\mathbb{E}\big[\mathrm{Var}\big(R_x(t)|R_x(s)\big)\big]$, the variability independent of $R_x(s)$. The variability $\mathrm{Var}\big(\mathbb{E}\big[R_x(t)|R_x(s)\big]\big)$ is called the shape component and $\mathbb{E}\big[\mathrm{Var}\big(R_x(t)|R_x(s)\big)\big]$ is called the magnitude

component. By analyzing the ratio of the shape component to the total variation depth, a definition of the shape variation is introduced.

Definition 3.33. For any function $x \in \mathcal{C}(U)$, the shape variation (SV) of x with respect to the distribution P is defined by

$$\mathrm{SV}(x) = \int_U v\big(t, s(t)\big) S_x\big(t, s(t)\big) dt,$$

where $s(t)$ is chosen such that for each $t \in U$, $s(t) \leq t$, where $v\big(t, s(t)\big)$ is a weight function given by

$$v\big(t, s(t)\big) := \frac{\big|x(t) - x\big(s(t)\big)\big|}{\int_U \big|x(t) - x\big(s(t)\big)\big|},$$

and $S_x\big(t, s(t)\big)$ is defined by

$$S_x\big(t, s(t)\big) := \begin{cases} \dfrac{\mathrm{Var}(\mathbb{E}[R_x(t)|R_x(s)]}{D_{S1}(x(t), P_t)} & , \text{ for } D_{S1}\big(x(t), P_t\big) \neq 0 \\ 1 & , \text{ for } D_{S1}\big(x(t), P_t\big) = 0 \end{cases}.$$

This shape variation can be used to successfully identify shape outliers, i.e. functions that may lie "in the middle" of the sample, but have a different shape than most of their peers.

Remark 3.34. To better reflect the outlyingness of a function x that may have an overall small depth $D_{S1}\big(x(t), P_t\big)$, a modified shape variation is also proposed in Huang and Sun (2016).

3.3 Depth Functions For Multivariate Functional Data

We now study K-variate stochastic processes $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(K)}) \in \mathcal{C}(U)^K$ and keep using notation from Subsection 2.2.2.

3.3.1 Multivariate Integrated Depth

The *multivariate integrated depth*[12] (Claeskens et al., 2014) is a generalization of the integrated depth in (3.20) for the K-variate functional space.

Definition 3.35. Let D_M be any depth for multivariate data. Then the multivariate integrated depth of $\boldsymbol{x} \in \mathcal{C}(U)^K$ with respect to P is defined by

$$D_{MID}(\boldsymbol{x}, P, D_M) = \int_U D_M\big(\boldsymbol{x}(t), P_t\big) w(t) dt, \tag{3.36}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^K$ is the K-variate function $\boldsymbol{x}$ evaluated at $t \in U$ and $w(\cdot)\colon U \longmapsto \mathbb{R}^+$ is a non-negative weight function.

[12] Claeskens et al. (2014) introduced this depth under the name "multivariate functional depth", but given its nature, we found the name "multivariate integrated depth" to be more appropriate.

One possibility for the weight function w is to choose a constant function over an interval of interest. However, to take into account the variability of the data, the following weight function is proposed in Claeskens et al. (2014)

$$w_3(t) = w_3(t, P_t, D_M) = \text{vol}\{C_\alpha(P_t, D_M)\} / \int_U \text{vol}\{C_\alpha(P_u, D_M)\} du, \tag{3.37}$$

where $C_\alpha(P_t, D_M)$ is the *marginal depth central region* at level $\alpha \geq 0$ with respect to P_t, defined by

$$C_\alpha(P_t, D_M) := \{\boldsymbol{x} \in \mathbb{R}^K : D_M(\boldsymbol{x}, P_t) > \alpha\}. \tag{3.38}$$

If w_3 is used, we denote the corresponding multivariate integrated depth by $D_{MID,\alpha}$.

To obtain the empirical version of (3.36) replace, as usual, P and P_t by, respectively, P_n and $P_{t,n}$.

Remark 3.36. Note that in many cases the definition of $D_{MID,\alpha}$ is independent of the choice of α since the depth regions are often proportional to a fixed function of α. In particular, for many depth functions on multivariate data and for unimodal elliptic symmetric distributions, the contours of the depth regions coincide with density contours.

Remark 3.37. Claeskens et al. (2014) examined in detail the multivariate integrated depth with the Tukey depth as the marginal depth D_M. They called the resulting depth the "multivariate half-space depth".

Remark 3.38. One possible application of the multivariate integrated depth is to smooth functional data and their derivatives, an application that is studied in Claeskens et al. (2014) and in Hlubinka, Gijbels, Omelka, and Nagy (2015).

3.3.2 Multivariate Weighted Integrated Depth

The *multivariate weighted integrated depth*[13] (Hlubinka et al., 2015) is similar to the multivariate integrated depth in (3.36).

Definition 3.39. Let D_M be any depth for univariate data, then the multivariate weighted integrated depth of $\boldsymbol{x} \in \mathcal{C}(U)^K$ with respect to P is defined by

$$D_{MWID}(\boldsymbol{x}, P, D_M) = \sum_{i=1}^{K} w_i \int_U D_M\big(x^{(i)}(t), P_t\big) dt, \tag{3.39}$$

where $x^{(i)}(t)$ is the i-th marginal function of $\boldsymbol{x}$ evaluated at $t \in U$ and $w_i > 0$, $i = 1, \ldots, K$, are non-negative weights such that $\sum_{i=1}^{K} w_i = 1$.

Here, the integrated depth defined in (3.20) is applied to each marginal function of $\boldsymbol{x}$ to obtain K separate depth values. These are then aggregated with weights to obtain the final depth value of $\boldsymbol{x}$.

[13] Since no name of this depth is given in (Hlubinka et al., 2015), this is the one we choose.

Remark 3.40. A similar approach to the multivariate weighted integrated depth is used in Ieva and Paganoni (2013) to obtain two other multivariate functional depth functions. Here, either the band depth defined in (3.27) or the modified band depth defined in (3.29) are applied to the marginal functions $x^{(i)}$, $i = 1, \ldots, K$, instead of the integrated depth in (3.39).

3.3.3 Multivariate Functional Random Projection Depth

The *multivariate functional random projection depth*[14] (Cuevas et al., 2007) is the multivariate extension of the functional random projection depth defined in (3.21). We here too use k "random projection angle functions" v_i, $i = 1, \ldots, k$, and this time apply these to each marginal function of the K-variate function $\boldsymbol{x}$ to obtain a vector in $\mathbb{R}^K$ for each $i = 1, \ldots, k$. These vectors are then analyzed with a depth function of choice for multivariate data to obtain the final depth value for $\boldsymbol{x}$. Here, we need $x_1, \ldots, x_n \in \mathcal{F}_K := \mathcal{L}^2(U)^K \cap \mathcal{C}(U)^K$, where $\mathcal{L}^2(U)^K$ is the K-variate space of square integrable functions on U.

Definition 3.41. Let D_M be any depth for multivariate data and let η be a non-degenerate probability distribution on $\mathcal{C}(U) \cap \mathcal{L}^2(U)$. The multivariate functional random projection depth of $\boldsymbol{x} \in \mathcal{F}_K$ with respect to P based on k stochastic processes with distribution η is

$$D_{MFRPD,k,\eta}(x, P, D_M) := \frac{1}{k} \sum_{i=1}^{k} D_M\big(\langle v_i, \boldsymbol{x} \rangle, P_{v_i}\big), \tag{3.40}$$

where

$$\langle v_i, \boldsymbol{x} \rangle := \Big(\big\langle v_i, x^{(1)} \big\rangle, \ldots, \big\langle v_i, x^{(K)} \big\rangle\Big)$$

is the vector in of component-wise inner products between v_i and $x^{(j)}$ defined in (3.22), $i = 1, \ldots, k$, $j = 1, \ldots, K$ and P_{v_i} denotes the K-variate marginal distribution of P on $\{\langle v_i, \boldsymbol{y} \rangle : \boldsymbol{y} \in \mathcal{F}_K\}$.

Remark 3.42. The distribution η is typically chosen to be a non-degenerate stationary Gaussian process on $\mathcal{C}(U) \cap \mathcal{L}^2(U)$.

To get the sample analogue of (3.40) we replace, as usual, P by P_n and P_{v_i} by $P_{v_i,n}$.

Remark 3.43. Two particular multivariate functional random projection depths cases are considered in Cuevas et al. (2007). The first one, called the *double random projection depth* uses $D_M = D_{T,k',\nu}$, i.e. the random Tukey depth with k' random projections defined in (3.6). The second one, called the *random projection depth*, uses D_M equal to the h-depth, a local depth for multivariate data also introduced in Cuevas et al. (2007). These two depth functions are then applied to bivariate functional data: a random sample of univariate functions $x_1, \ldots, x_n$ and their corresponding derivatives.

Remark 3.44. Similarly to functional random projection depth defined in (3.21), the multivariate random projection depth can also be modified by replacing the mean in (3.40) by the minimum over

[14] No formal definition of this depth is given in (Cuevas et al., 2007), so we have generalized it and expanded its name accordingly.

$i = 1, \ldots, k$. For later examinations, we refer to the two depth functions as $D_{MFRPD,k,\eta,\text{mean}}(x, P, D_M)$ and $D_{MFRPD,k,\eta,\min}(x, P, D_M)$ if the distinction is required.

Remark 3.45. Since the projections for the multivariate functional random projection depth are applied to the marginals $x^{(i)}, i = 1, \ldots, K$, we use the same number k_0 of "functional" projections determined in the univariate functional case presented in Section 3.2.2, i.e. a conservative value $k_0 = 1\,000$. Note that if $D_M = D_{T,k',\nu}$ is used, the number of required multivariate projections k_0' needs to be determined with respect to the dimension K using the results from Section 3.1.2.

3.3.4 Simplicial Band Depth

The *simplicial band depth* (López-Pintado et al., 2014) is the generalization to the K-variate functional case of the band depth introduced in (3.27) with $J = 2$.

Definition 3.46. The simplicial band depth of $\boldsymbol{x} \in \mathcal{C}(U)^K$ with respect to P is

$$D_{SBD}(\boldsymbol{x}, P) := P\Big\{\boldsymbol{x} \in SB\big[\boldsymbol{X_1}(t), \ldots, \boldsymbol{X_{K+1}}\big]\Big\}, \tag{3.41}$$

where $\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}$ are independent stochastic processes with distribution P and $SB\big[\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}\big]$ denotes the random "simplex band" with "vertices" $\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}$ defined by

$$SB\big[\boldsymbol{X_1}, \ldots, \boldsymbol{X_{K+1}}\big] := \Big\{\boldsymbol{Y} : \boldsymbol{Y}(t) = \alpha_1 \boldsymbol{X_1}(t) + \cdots + \alpha_{K+1} \boldsymbol{X_{K+1}}(t) \quad \forall t \in U : \sum_{i=1}^{K+1} \alpha_i = 1 : \alpha_i > 0 \; \forall i\Big\}. \tag{3.42}$$

Analogously to the band depth in (3.27), to obtain the empirical version of (3.41), we use the $\binom{n}{K+1}$ simplex bands generated by the curves $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$ and come to the following definition.

Definition 3.47. The empirical simplicial band depth of $\boldsymbol{x} \in \mathcal{C}(U)^K$ with respect to P_n is

$$D_{SBD}(\boldsymbol{x}, P_n) := \binom{n}{K+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{K+1} \le n} \mathbb{1}\Big\{\boldsymbol{x} \in SB\big[\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}\big]\Big\},$$

where $SB\big[\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}\big]$ is the "simplex band" with "vertices" $\boldsymbol{x_{i_1}}, \ldots, \boldsymbol{x_{i_{K+1}}}$ defined by

$$SB\big[\boldsymbol{x_1}, \ldots, \boldsymbol{x_{K+1}}\big] := \Big\{\boldsymbol{y} \in \mathcal{C}(U)^K : \boldsymbol{y}(t) = \alpha_1 \boldsymbol{x_1}(t) + \cdots + \alpha_{x+1} \boldsymbol{x_{K+1}}(t) \quad \forall t \in U : \sum_{i=1}^{K+1} \alpha_i = 1 : \alpha_i > 0 \; \forall i\Big\}. \tag{3.43}$$

Remark 3.48. A modified version of (3.41) called the *modified simplicial depth* (that we denote by D_{MSBD}) is also introduced in López-Pintado et al. (2014). However, we skip its formal introduction as the modification step is identical to the one used for the modified band depth defined in (3.29).

Additionally, it is shown in López-Pintado et al. (2014), that the modified simplicial depth is equal to the multivariate integrated depth introduced in (3.36) with $D_M = D_S$, i.e. the simplicial depth defined in (3.15).

4 Comparison

4.1 Functional Data Depth Overview

As already noted, several depth functions for functional data are closely related to one another. As a result, we have decided to organize them into three groups.

The first group consists of depth functions that are based on integrals of marginal depth functions. Here, a depth function for univariate or multivariate data is applied to the marginals of the data $\{\boldsymbol{x_1}(t), \ldots, \boldsymbol{x_n}(t)\}$, $n \in \mathbb{N}$, for all $t \in U$ and subsequently integrated over t. One element of this group is the multivariate integrated depth, for which we can directly identify two special cases: the integrated depth and the total variation depth. Further, it is shown in Nagy, Gijbels, Omelka, and Hlubinka (2016), that the modified half-region depth, the modified band depth and the modified simplicial band depth are also special cases of the multivariate integrated depth. Finally, we also add to this group the weighted integrated depth and the extremal depth.

The second group consists of depth functions that are based on random projections. Here, functional random projections are used to reduce the original K-variate functional data to K-variate vectors for $K \in \mathbb{N}$, on which a data depth for univariate or multivariate data is applied. In this group, we have the multivariate functional random projection depth and the corresponding one dimensional version, the functional random projection depth. These include both possible modifications with the mean and the minimum.

For the third group, we consider geometric depth functions that rely on comparisons of the entire functions with one another. In this group we find the half-region depth, the band depth and the simplicial band depth. Note that in Mosler and Polyakova (2012), a more theoretical framework for this group is introduced with the development of so called Φ-*depths* in a Banach space.

Additionally, concerning the choice of the required marginal depth function in the first two groups, we notice a particular preference in the literature to often choose between the Tukey depth and the simplicial depth, as is the case for the latter in the modified band depth for $J = 2$, the total variation depth and the modified simplicial band depth. The Tukey depth, on the other hand, is used in the extremal depth and in the functional random Tukey depth.

For an overview, we visualize all the mentioned relationships between depth functions in Figure 4.1. Note that a summary of all used abbreviations is found in Chapter 7.

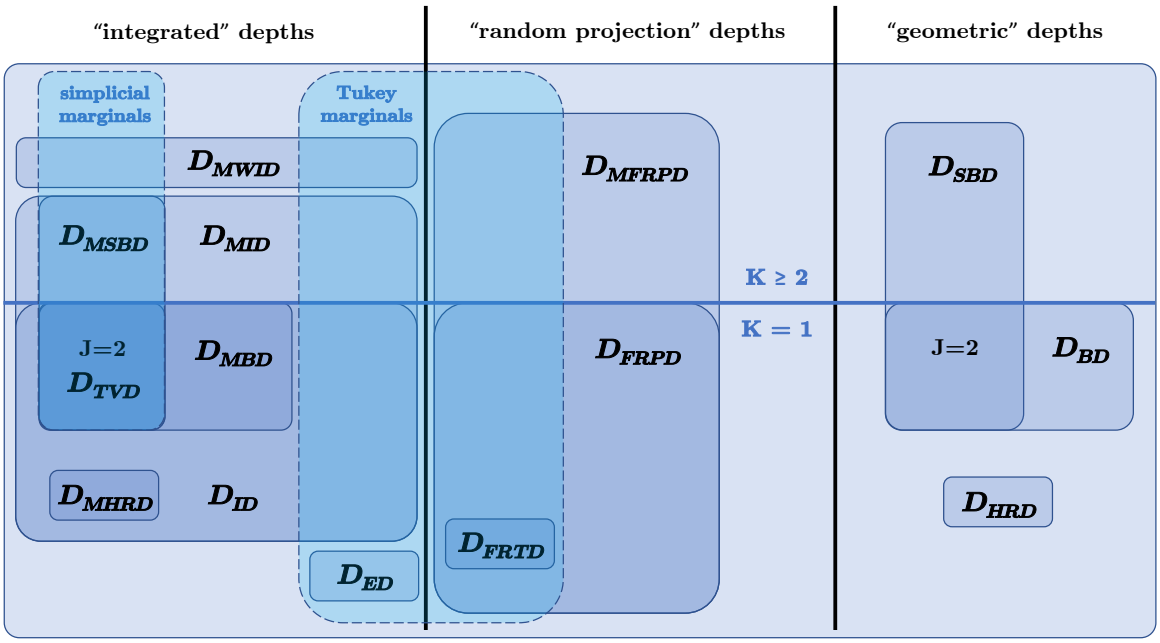


Figure 4.1: Relationship diagram between presented functional depths. The terms "simplicial" and "Tukey marginals" have been used to describe the choice for the marginal depth in each group. Note that the weighted multivariate integrated depth can also have univariate integrated depths as special cases; this was left out in this overview to simplify the plot.

4.2 Property-wise Analysis of Functional Depths

We now assess the performance of the introduced functional depth functions and begin with the analysis of Properties P1 to P4 presented in Definition 2.2 in the following theorem.

Consider the functional depth functions introduced in Chapter 3. For all functional depth functions requiring a marginal depth D_M for univariate or multivariate data, assume D_M to satisfy Properties M1 to M4 introduced in 2.1. Then the results in Table 4.1 hold.

	P1	P2	P3	P4
Integrated depths				
Multivariate integrated depth	✓	✓	✓	✗
Integrated depth	✓	✓	✓	✗
Modified simplicial band depth	✓	✓	✓	✗
Modifieid band depth	✓	✓	✓	✗
Total variation depth	✓	✓	✓	✗
Modified half-region depth	✓	✓	✓	✗
Multivariate weighted integrated depth	✓	✓	✓	✗
Extremal depth	✓	✓	✓	✓
Random projections depths				

FRPD with mean	✓	✓	✓	✓
FRPD with minimum	✓	✓	✓	✓
Multivariate FRPD with mean	✓	✓	✓	✓
Multivariate FRPD with minimum	✓	✓	✓	✓
Geometric depths				
Simplicial band depth	✓	✓	✓	✓
Band depth	✓	✓	?	✓
Half-region depth	✓	✗	?	✓

Table 4.1: Summary of properties for introduced functional depth functions. The symbol ✓ indicates that the property is satisfied and the symbol ✗ indicates that it is not. The question mark indicates that it is unknown if the property is satisfied and FRPD stands for "functional random projection depth".

Proof. Proofs for Property P1 to Property P3 for the multivariate integrated depth are found in Nagy et al. (2016). Those proofs are also applicable to the integrated depth, the modified simplicial band depth, the modified band depth, the total variation depth, the modified half-region depth and the multivariate weighted integrated depth as it is also shown in Nagy et al. (2016) that these are special cases of the multivariate integrated depth. In Narisetty and Nair (2015), it is shown that Property P4 does not hold for the integrated depth and thus, it also fails to hold for all the other depth functions named above.

Property P1 to Property P4 for the extremal depth are proven in Narisetty and Nair (2015).

Property P1 to Property P4 for the simplicial band depth are proven in López-Pintado et al. (2014) and thus also for its special case the band depth for $J = 2$. For $J \neq 2$, Property P1, Property P2 and Property P4 are proven in López-Pintado and Romo (2009). Property P3 is not addressed here.

Property P1 and Property P4 for the half-region depth are proven in López-Pintado and Romo (2011) and it is shown in Nieto-Reyes and Battey (2014) that property P2 does not hold. We did not find any proof for Property P3.

We now come to the proofs for the univariate and multivariate functional random projection depth for both the mean and minimum versions. For this, assume $\boldsymbol{x}$ to be element of $\mathcal{F}_K = \mathcal{C}(U)^K \cap \mathcal{L}^2(U)^K$, as required in the definition of the multivariate functional random projection provided in (**??**).

Property P1 - *Affine invariance*. We need to show that

$$D_{MFRPD,k,\eta,\text{mean}}(\boldsymbol{A}\boldsymbol{x}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}, P_{\boldsymbol{A}\boldsymbol{X}_{(eta+d)}+\tilde{\boldsymbol{X}}_{(ct+d)}}) = D_{MFRPD,k,\mu,\text{mean}}(\boldsymbol{x}, P_{\boldsymbol{X}})$$

holds for any constants $c \in \mathbb{R} \setminus \{0\}$, $d \in \mathbb{R}$, any vector of functions $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathcal{C}(U)^K$ with the interval $U := [l, u]$, any $P_{\boldsymbol{X}}, P_{\tilde{\boldsymbol{X}}} \in \mathcal{P}$ and any non-singular matrix $\boldsymbol{A} \in \mathbb{R}^{K\times K}$. Further, $\boldsymbol{A}\boldsymbol{x}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}$ is the function $s \longmapsto \boldsymbol{A}\boldsymbol{x}(\frac{s-d}{c}) + \tilde{\boldsymbol{x}}(\frac{s-d}{c})$ with $s \in S = [cl+d, cu+d]$, $\tilde{\boldsymbol{x}}_{(ct+d)}$ is the function $s \longmapsto \tilde{\boldsymbol{x}}(\frac{s-d}{c})$, $s \in S$, $\boldsymbol{A}\boldsymbol{X}_{(ct+d)} + \tilde{\boldsymbol{X}}_{(ct+d)}$ is the stochastic process $\{\boldsymbol{A}\boldsymbol{X}(\frac{s-d}{c}) + \tilde{\boldsymbol{X}}(\frac{s-d}{c}) : s \in S\}$ and $\tilde{\boldsymbol{X}}_{(ct+d)}$ is the

stochastic process $\{\tilde{\boldsymbol{X}}(\frac{s-d}{c}) : s \in S\}$.

Consider for any $v_i \in \{v_1, \ldots, v_k\}$,

$$\begin{aligned}\langle \boldsymbol{A}\boldsymbol{x}_{(ct+d)} + \tilde{\boldsymbol{x}}_{(ct+d)}, v_{i,(ct+d)} \rangle &= \int_{S=[cl+d,cu+d]} \Big[\boldsymbol{A}\boldsymbol{x}\big(\tfrac{s-d}{c}\big) + \tilde{\boldsymbol{x}}\big(\tfrac{s-d}{c}\big)\Big] v_i\big(\tfrac{s-d}{c}\big) ds \\ &= \int_U \Big[\boldsymbol{A}\boldsymbol{x}(t) + \tilde{\boldsymbol{x}}(t)\Big] v_i(t) dt \quad \text{with } t = \tfrac{s-d}{c} \\ &= \boldsymbol{A}\langle \boldsymbol{x}, v_i \rangle + \langle \tilde{\boldsymbol{x}}, v_i \rangle.\end{aligned}$$

Since $\langle \boldsymbol{x}, v_i \rangle$ and $\langle \tilde{\boldsymbol{x}}, v_i \rangle$ are values in $\mathbb{R}^K$ for any $v_i \in \{v_1, \ldots, v_k\}$, Property P1 follows from D_M satisfying Property M1. Also, since the above equations hold before applying the mean or the minimum, Property P1 holds for both versions of the multivariate functional random projection depth.

Property P2 - *Maximality at the center.* We need to show that

$$D_{MFRPD,k,\eta,\text{mean}}(\boldsymbol{\Theta}, P_{\boldsymbol{X}}) = \sup_{\boldsymbol{x} \in \mathcal{F}_K} D_{MFRPD,k,\eta,\text{mean}}(\boldsymbol{x}, P_{\boldsymbol{X}})$$

holds for any $P_{\boldsymbol{X}} \in \mathcal{P}$ having a unique center of functional symmetry $\boldsymbol{\Theta} \in \mathcal{F}_K$.

For any $v_i \in \{v_1, \ldots, v_k\}$, we have that $\langle \mathbb{E}\boldsymbol{X}, v_i \rangle$ is the mean of P_{v_i} because

$$\begin{aligned}\mathbb{E}\big[\langle \boldsymbol{X}, \boldsymbol{v_i} \rangle\big] &= \mathbb{E}\left[\int_U \boldsymbol{X}(t) v_i(t) dt\right] \\ &= \int_U \mathbb{E}\boldsymbol{X}(t) v_i(t) dt,\end{aligned}$$

where the last step holds by applying Fubini's Theorem (Fubini, 1907). Since the distribution of a Gaussian process is a symmetric, we have that $\langle \mathbb{E}\boldsymbol{X}, v_i \rangle$ is the center of symmetry of P_{v_i} and thus Property P2 follows from D_M satisfying Property M2. This holds again for both the mean and minimum versions for the same previous reasons.

Property P3 - *Monotonicity relative to the deepest point.* Here, we need to show that for any $P_{\boldsymbol{X}} \in \mathcal{P}$ having a deepest point $\boldsymbol{\Theta}$,

$$D_{MFRPD,k,\eta,\text{mean}}(\boldsymbol{x}, P_{\boldsymbol{X}}) \leq D_{MFRPD,k,\eta,\text{mean}}\big(\boldsymbol{\Theta} + \alpha(\boldsymbol{x} - \boldsymbol{\Theta}), P_{\boldsymbol{X}}\big)$$

holds for any $\boldsymbol{x} \in \mathcal{F}_K$ and for any $\alpha \in [0, 1]$.

Since $\langle \mathbb{E}\boldsymbol{X}, v_i \rangle$ is the maximum of the distribution P_{v_i}, as shown in the proof of Property P2, Property P3 follows from D_M satisfying Property M3, again also for both versions.

Property P4 - *Null at the boundary.* We need to show that

$$\lim_{m \to \infty} D_{MFRPD,k,\eta,\text{mean}}(\boldsymbol{x_m}, P_{\boldsymbol{X}}) = 0$$

for any $P_{\boldsymbol{X}} \in \mathcal{P}$ and for all sequences of functions $\boldsymbol{x_m} \in \mathcal{F}_K$ with $\lim_{m\to\infty} |x_m^{(j)}(t)| = \infty$ for at least one $t \in U$ and one $j \in \{1, \ldots, K\}$.

Assume $x_m^{(j)} \in \mathcal{C}(U) \cap \mathcal{L}^2(U)$ with $\lim_{m\to\infty} |x_m^{(j)}(t)| = \infty$ for at least one $t \in U$ and one $j \in \{1, \ldots, K\}$. Then, because $x_m^{(j)} \in \mathcal{C}(U)$, there exist an $\epsilon > 0$ such that $\lim_{m\to\infty} |x_m^{(j)}(s)| = \infty$ for all $s \in [t-\epsilon, t+\epsilon]$. Then, for any $v_i \in \{v_1, \ldots, v_k\}$

$$\begin{aligned}
&\lim_{m\to\infty} \left|\langle x_m^{(j)}, v_i\rangle\right| \\
&= \lim_{m\to\infty} \left| \int_U x_m^{(j)}(s) v_i(s) ds \right| \\
&= \lim_{m\to\infty} \left| \int_{U\setminus[t-\epsilon,t+\epsilon]} x_m^{(j)}(s) v_i(s) ds + \int_{[t-\epsilon,t+\epsilon]} x_m^{(j)}(s) v_i(s) ds \right|. \qquad (4.1)
\end{aligned}$$

Since $v_i \in \mathcal{C}(U)$, we assume that ϵ is small enough so that v_i does not change sign over the interval $[t-\epsilon, t+\epsilon]$. Thus, the last term in (4.1) goes to $\pm\infty$ and thus $\lim_{m\to\infty} \left|\langle x_m^{(j)}, v_i\rangle\right| = \infty$. Now Property P4 follows again because Property M4 is satisfied for D_M and again for both versions of the multivariate functional random projection depth.

Finally, the functional random projection depth satisfies Property P1 to Property P4 since it is a special case of the multivariate functional random projection depth.

□

From Table 4.1, we notice that almost all depth functions satisfy Property P1 to Property P3, with the exception of the half-region depth, where Property P2, i.e. the *maximality at center* property, is not satisfied. Possibly Property P3, i.e. *the monotonicity relative to the deepest point* property, might also not be satisfied by the half-region depth and the band depth.

However, all three introduced geometric depth functions suffer from degeneracy, making the four studied properties irrelevant for these depth functions. For instance, in Chakraborty and Chaudhuri (2014), it is shown that for many commonly used stochastic process models, the half-region depth and the band depth lead to all realizations having zero depths with probability one. That is, because of many crossings between functions, no single function lies within a band or a half-region. A fortiori, the simplicial band depth also suffers from this degeneracy and we thus stop considering any geometric depth going forward.

Property P4, on the other hand, is not satisfied by the multivariate integrated depth and its special cases. This can easily be seen by considering a function $x \in \mathcal{C}(U)$ that has values outside of the convex hull of the data on only some interval $[a, b] \in U$ and then considering the integral in the multivariate integrated depth. (For this, recall the visualization of the integrated depth in Figure 3.7.) The extremal depth, because it priorities "extremes" in its construction, leads to Property P4 being satisfied. So do all versions of the univariate and multivariate functional random projections, as just shown.

To further differentiate between the presented functional depth functions, we study some additional

properties and features. In the multivariate weighted integrated depth introduced in (3.39), a univariate functional depth function is applied to K marginal functions of $\boldsymbol{x} \in \mathcal{C}(U)^K$, $K \in \mathbb{N}$. The resulting depth values are then aggregated using arbitrary weights to obtain the final depth value. However, by performing this depth calculation, any dependence structure between the marginal functions of $\boldsymbol{x}$ is ignored (c.f. (3.39)). This is not the case, for example, in the multivariate integrated depth defined in (3.36). While the depth is constructed in a similar way, it also incorporates the mentioned dependence structure and as a result, makes the multivariate weighted integrated depth obsolete.

In a simulation study in Narisetty and Nair (2015), desirable relationships between central regions of the extremal depth and their probability coverage are discovered. Additionally, another simulation study in the same paper reveals that the extremal depth performs better than the modified band depth and the integrated depth for outlier detection.

In a similar setting, it is suggested in Huang and Sun (2016), that the outlier detection based on the total variation depth and its decomposition is superior to the outlier detection based on the extremal depth. In particular, a robust differentiation between shape and magnitude outlyingness is achieved.

As a preliminary conclusion, by removing special cases and non-performing depth functions we narrow the studied functional depth functions down to the following four depth functions: the multivariate integrated depth, the extremal depth and the multivariate functional random projection depth with mean and with minimum. This implicitly includes the total variation depth, a special case of the multivariate integrated depth, for which the total variation decomposition might be of additional interest.

To finish this section, recall Property P5 introduced in Section 2.2. Since this property is informally defined, only conjectures about it can be made.

For the multivariate integrated depth, the weight function w_3 defined in (3.37) takes into account the variability of the underlying distribution over its range with the help of depth central regions calculated for all $t \in U$. As a result, the multivariate integrated depth would naturally satisfy property P5, if used with this weight function.

We suspect that Property P5 is also satisfied by the multivariate functional projection depth for both the mean and minimum version. As intuition we consider the following extreme case. Assume that for a sample of univariate functional data $\{x_1, \ldots, x_n\}$, $x_i \in \mathcal{C}(U)$, $i = 1, \ldots, n$, the variability is concentrated on the interval $[a, b] \subset U$, i.e. $x_1, \ldots x_n$ are equal for all $t \in U \backslash [a, b]$. Further, the inner products defined in (3.22) of the functional projections v_i, $i = 1, \ldots, k$ and the functions x_j, $j = 1, \ldots, n$ in are reformulated in the following way

$$\langle v_i, x_j \rangle = \int_U v_i(t) x_j(t) dt = \int_{[a,b]} v_i(t) x_j(t) dt + \int_{U \backslash [a,b]} v_i(t) x_j(t) dt.$$

Thus, given our assumptions, the terms $\int_{U \backslash [a,b]} v_i(t) x_j(t) dt$ are equal for all x_j, $j = 1, \ldots, n$ given a fixed v_i and as a result, all variability present in the distribution P_{v_i} comes from the terms $\int_{[a,b]} v_i(t) x_j(t) dt$, $j = 1, \ldots, n$ for any v_i, $i = 1, \ldots, k$. Since this holds before applying either the mean or the minimum function to the inner products, in this special case, Property P5 is satisfied for

both versions of the functional random projection depth.

Finally, the construction of the extremal depth does not suggest that it satisfies Property P5, since the univariate depth calculations of every t-marginals are computed independently from one another.

These intuitions and further depth characteristics are now studied in the following section, where we assess the performance of various depth functions visually.

4.3 Data Depth Visualization

We begin by simulating four data sets. The first two are from the same Gaussian process distributions used in various previous examples, e.g. in Figure 1.1 and in Figure 2.4. One data set has constant amplitude over the range U, while the other has high relative amplitude on a small interval. The simulation process is inspired by the Gaussian process regression, introduced in Ebden (2015), and explained in detail in the Appendix.

The other two data sets are generated from *max-stable* distributions, i.e. distributions that are known to produce "extreme" processes. One data set is from a Brown-Resnick process and the other is from an extremal-t process. Both can be simulated in the **R** package "SpatialExtremes" (Davison, Padoan, and Ribatet, 2012). For both processes we used a range of 3 and a smooth parameter of 0.7. For the extremal-t processes, we additionally set the number of degrees of freedom to 4. Note that these two data sets have constant amplitude over the range U. For more informations about these processes we refer to the coming Section 5.1.3 and the Appendix

In all four cases, we generate $n = 1\,000$ curves with $T = 100$ discretization points. The data is plotted in Figure 4.2.

We now examine this data with five different univariate integrated depths: the integrated depth with $D_M = D_{T1}$ we denote by D_{IDT1}, the integrated depth depth with $D_M = D_{S1}$ we denote by D_{IDS1}, the integrated depth with $D_M = D_{T1}$ and with weight function w_3 defined in (3.37), we denote this depth by $D_{IDT1,\alpha}$, the integrated depth with $D_M = D_{S1}$ and with weight function w_3, we denote this depth by $D_{IDS1,\alpha}$, and the modified half-region depth. We compare the resemblance between two depth functions by examining so called DD-plots and their corresponding correlation coefficient.

Definition 4.1. The *depth-versus-depth plots* (DD-plots) (Liu et al., 1999) is defined by

$$\mathrm{DD}(D_F, D_G, P_n) := \big\{\big(D_F(x_i, P_n), D_G(x_i, P_n)\big), i = 1, \dots, n\big\},$$

where $D_F(\,\cdot\,, P_n)$ and $D_G(\,\cdot\,, P_n)$ are two depth functions and $x_1, \dots, x_n$ are realizations of independent random variables with distribution P.

In a DD-plot the pair of points $\big(D_F(x_i, P_n), D_G(x_i, P_n)\big)$, $i = 1, \dots, n$ are plotted.

In the following figures, we plot a matrix of the pairwise DD-plots using normalized depths and we compute the corresponding correlation coefficients for each data set. On the diagonal, we visualize the data and color it by the corresponding depth values.

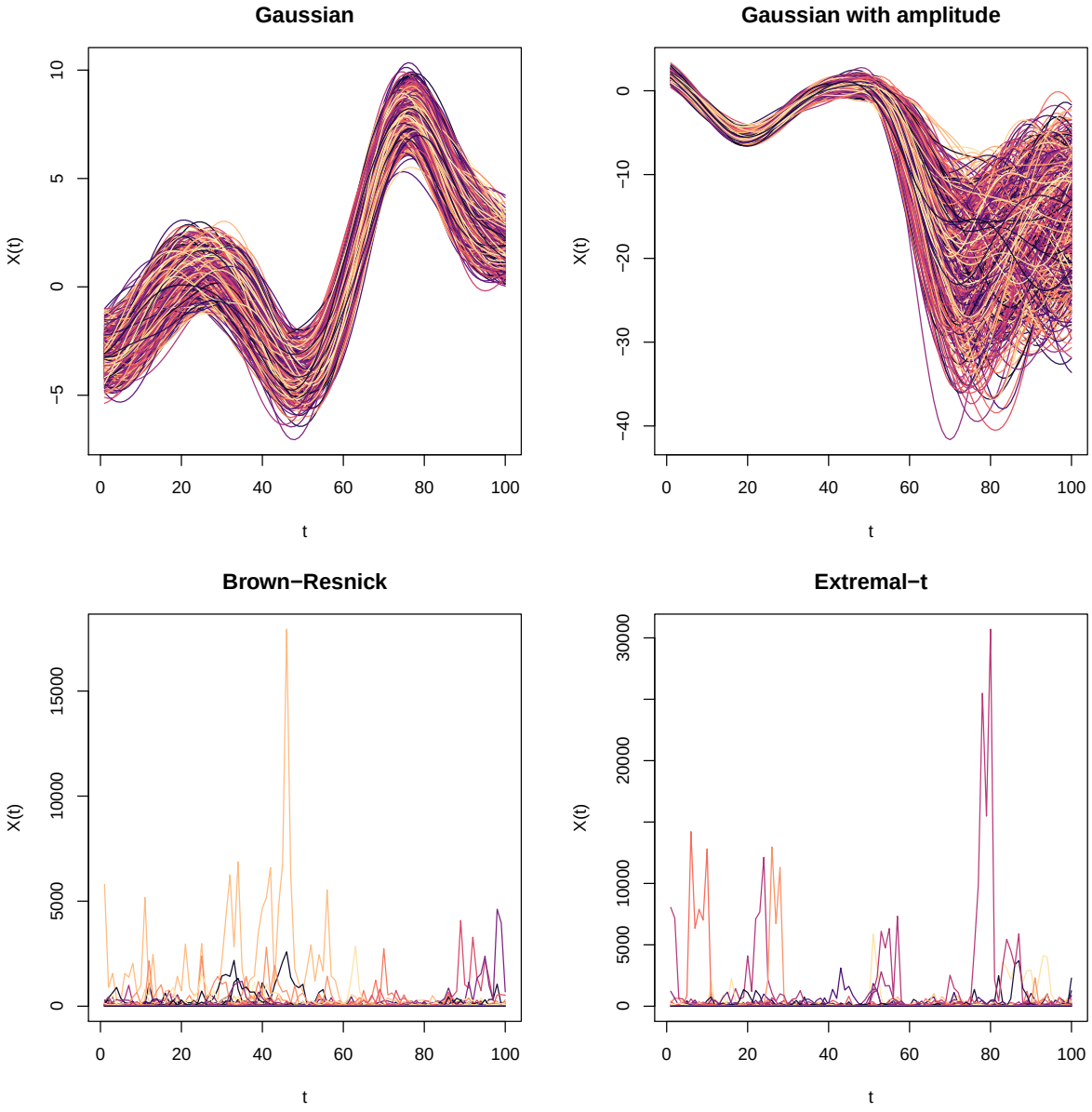


Figure 4.2: Four simulated data sets with $n = 1\,000$ curves and $T = 100$ discretization points. The first two data sets are drawn from Gaussian processes, while the last two are drawn from max-stable processes, one from a Brown-Resnick process the other from an extremal-t process. Colors are chosen randomly.

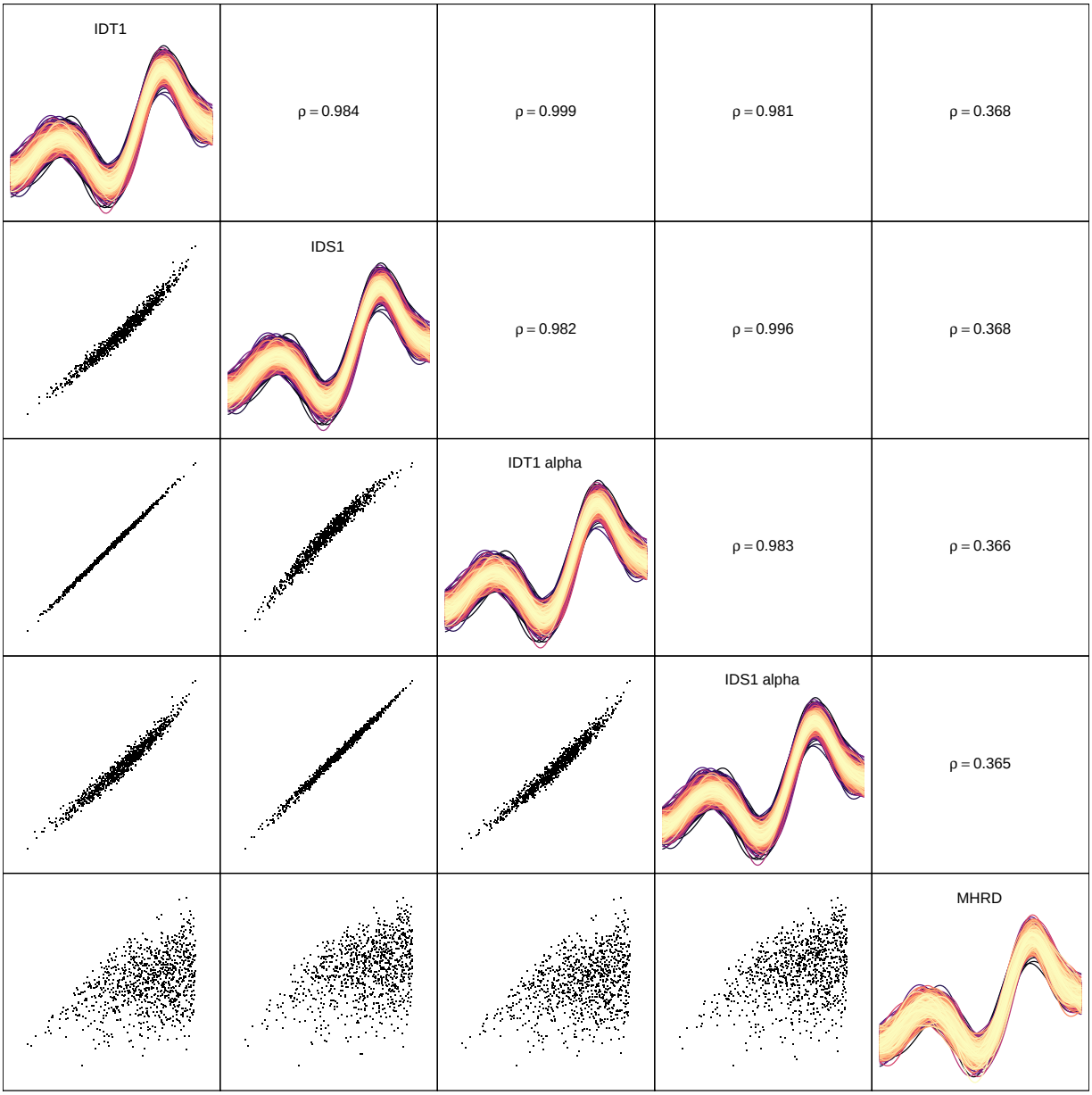


Figure 4.3: Matrix of the DD-plots for all pairs of D_{IDT1}, D_{IDS1}, $D_{IDT1,\alpha}$, $D_{IDS1,\alpha}$ and D_{MHR}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Gaussian process with constant amplitude.

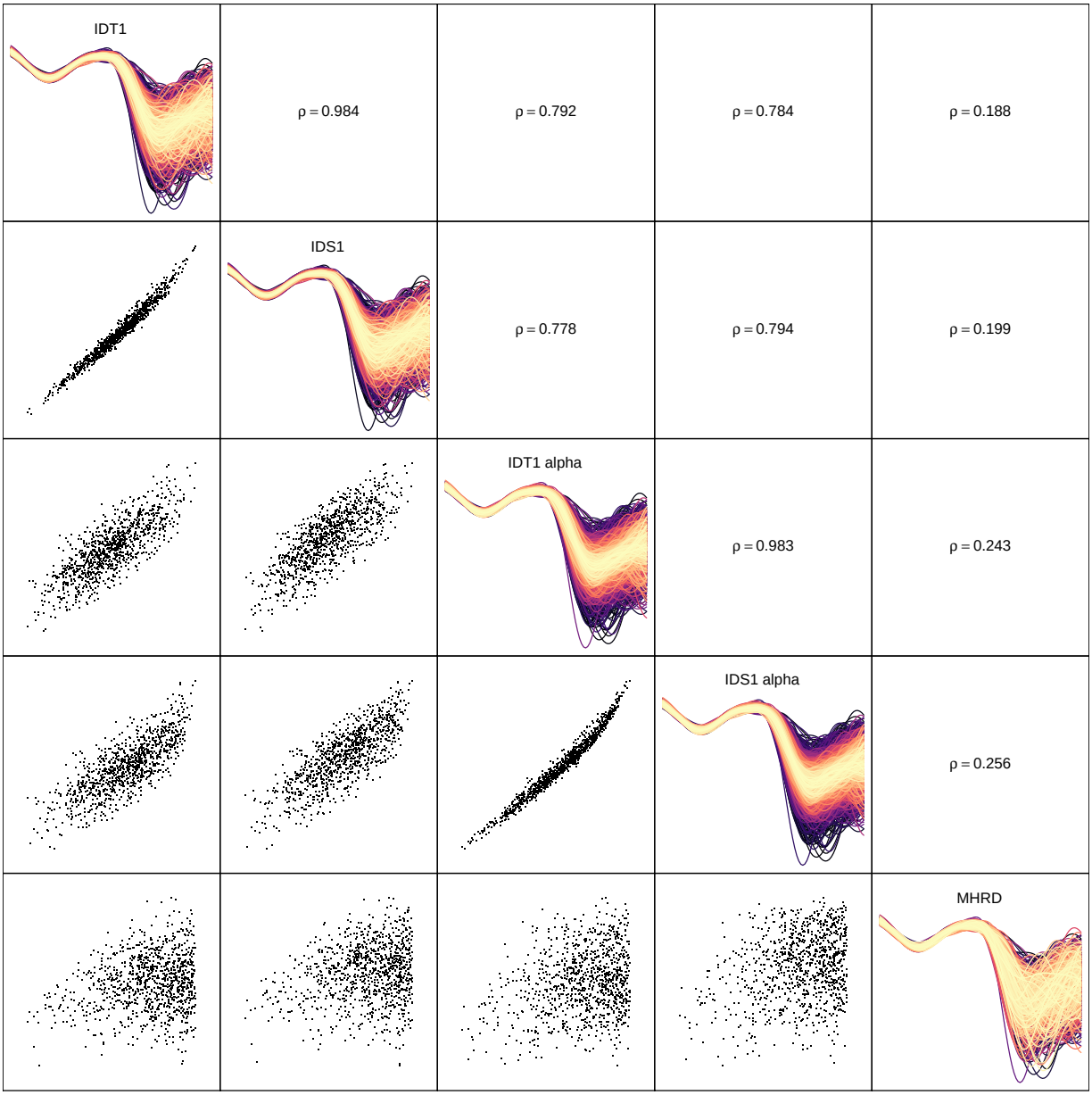


Figure 4.4: Matrix of the DD-plots for all pairs of D_{IDT1}, D_{IDS1}, $D_{IDT1,\alpha}$, $D_{IDS1,\alpha}$ and D_{MHR}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Gaussian process with varying amplitude.

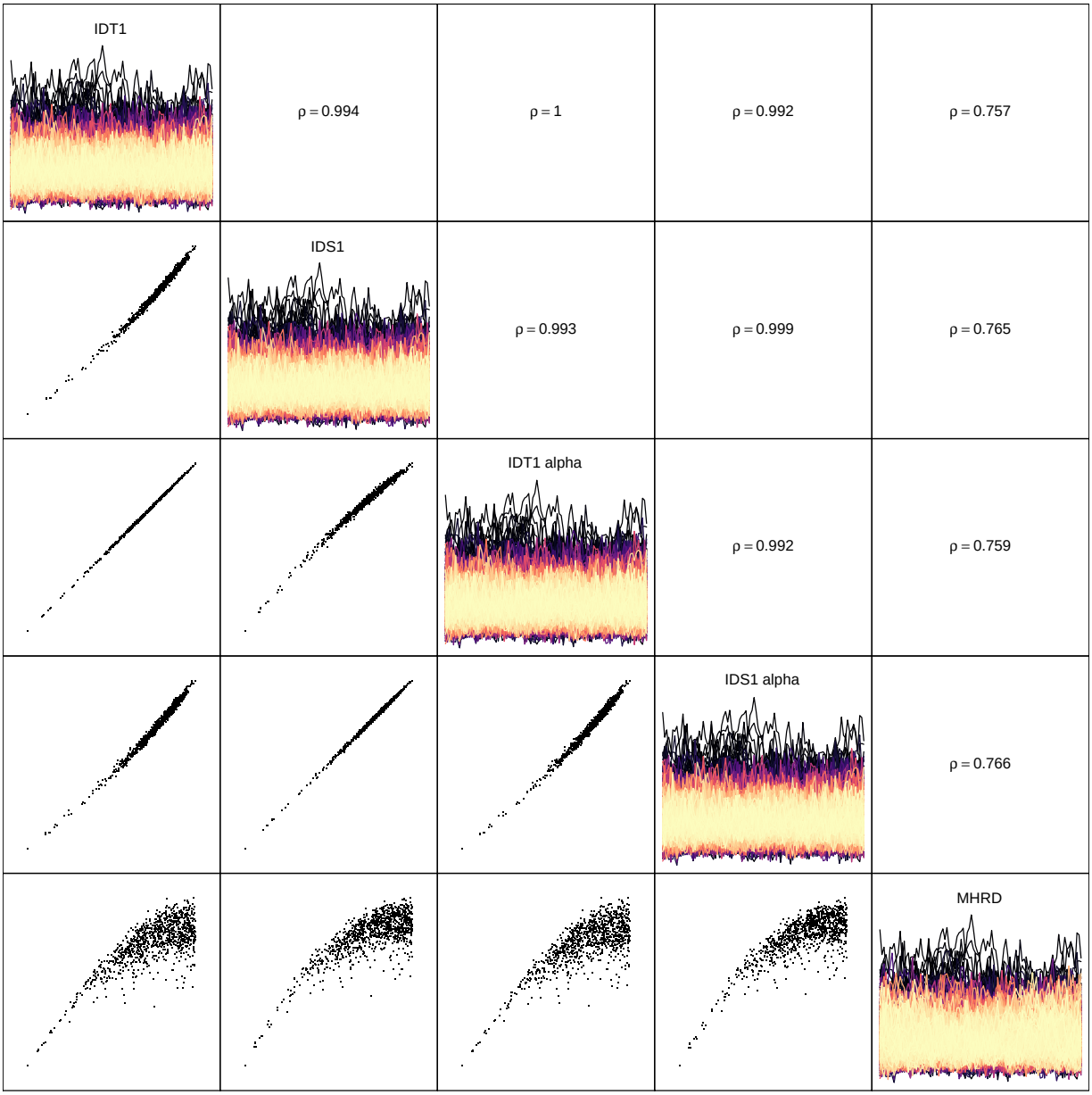


Figure 4.5: Matrix of the DD-plots for all pairs of D_{IDT1}, D_{IDS1}, $D_{IDT1,\alpha}$, $D_{IDS1,\alpha}$ and D_{MHR}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Brown-Resnick (logarithms).

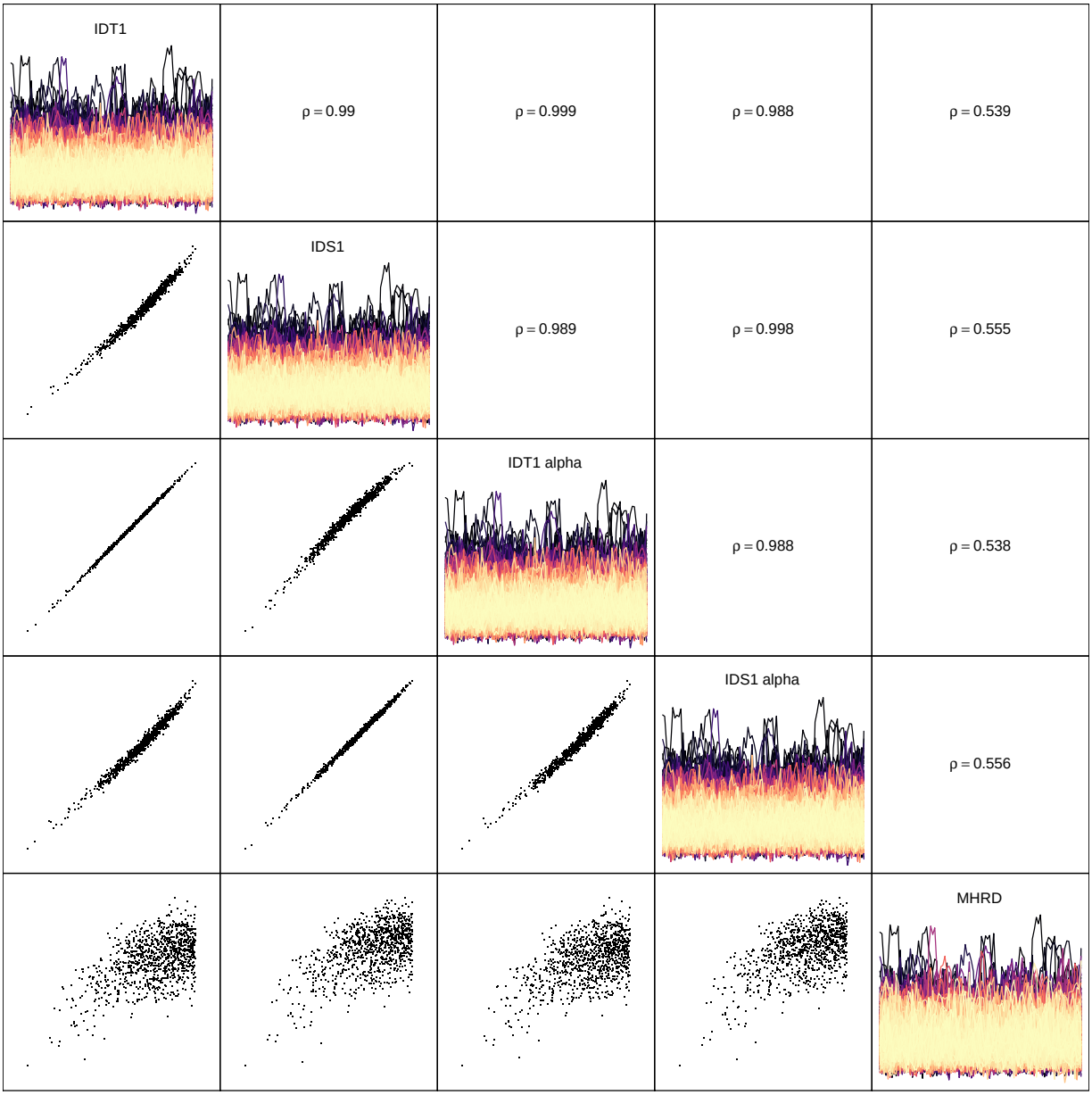


Figure 4.6: Matrix of the DD-plots for all pairs of D_{IDT1}, D_{IDS1}, $D_{IDT1,\alpha}$, $D_{IDS1,\alpha}$ and D_{MHR}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: extremal-t (logarithms)

For data with constant amplitude over the range U (i.e. data sets 1, 3 and 4), all integrated depths perform similarly, with the exception of the modified half-region depth. This is indicated by a high correlation and points in the DD-plot forming a straight line, i.e. the considered depth functions estimate similar depth values for each point. However, for the data set with varying amplitude, we observe differences. As expected from the last section, we notice that the integrated depths with weight function lead to narrower regions of functions colored in light colors on the right side of the plots indicating that they satisfy Property P5, while the other considered depth functions here do not. It also seems that the modified half-region depth fails to order each data set in any reliable way.

Taking for marginal depth the Tukey depth or the simplicial depth seems to be of relatively low importance. As a result, we only consider the integrated depth with and without weight function and with $D_M = D_{T1}$ when comparing the integrated depth to the functional random projection depths and the extremal depth in our next step. Here we denote the functional random projection depth with $D_M = D_{T1}$ by $D_{FRPDT1,\text{mean}}$ or $D_{FRPDT1,\text{min}}$ depending on which function is applied in (3.21). Additionally, the functional projections v_j are chosen from a Gaussian process in all depth calculations. We keep the same visualization framework as before and obtain the following plots.

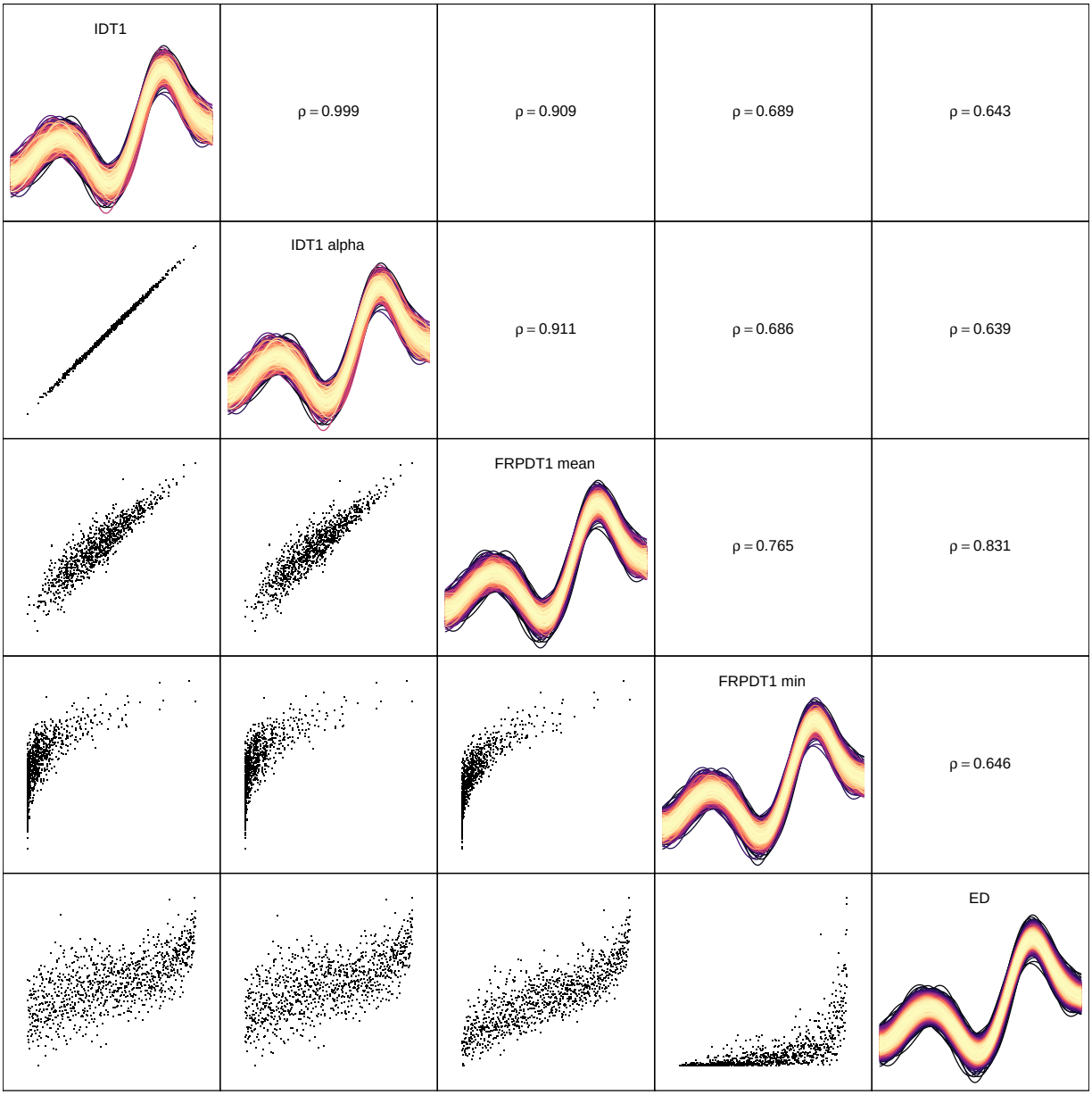


Figure 4.7: Matrix of the DD-plots for all pairs of D_{IDT1}, $D_{IDT1,\alpha}$, D_{FRPDT1} and D_{ED}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Gaussian process.

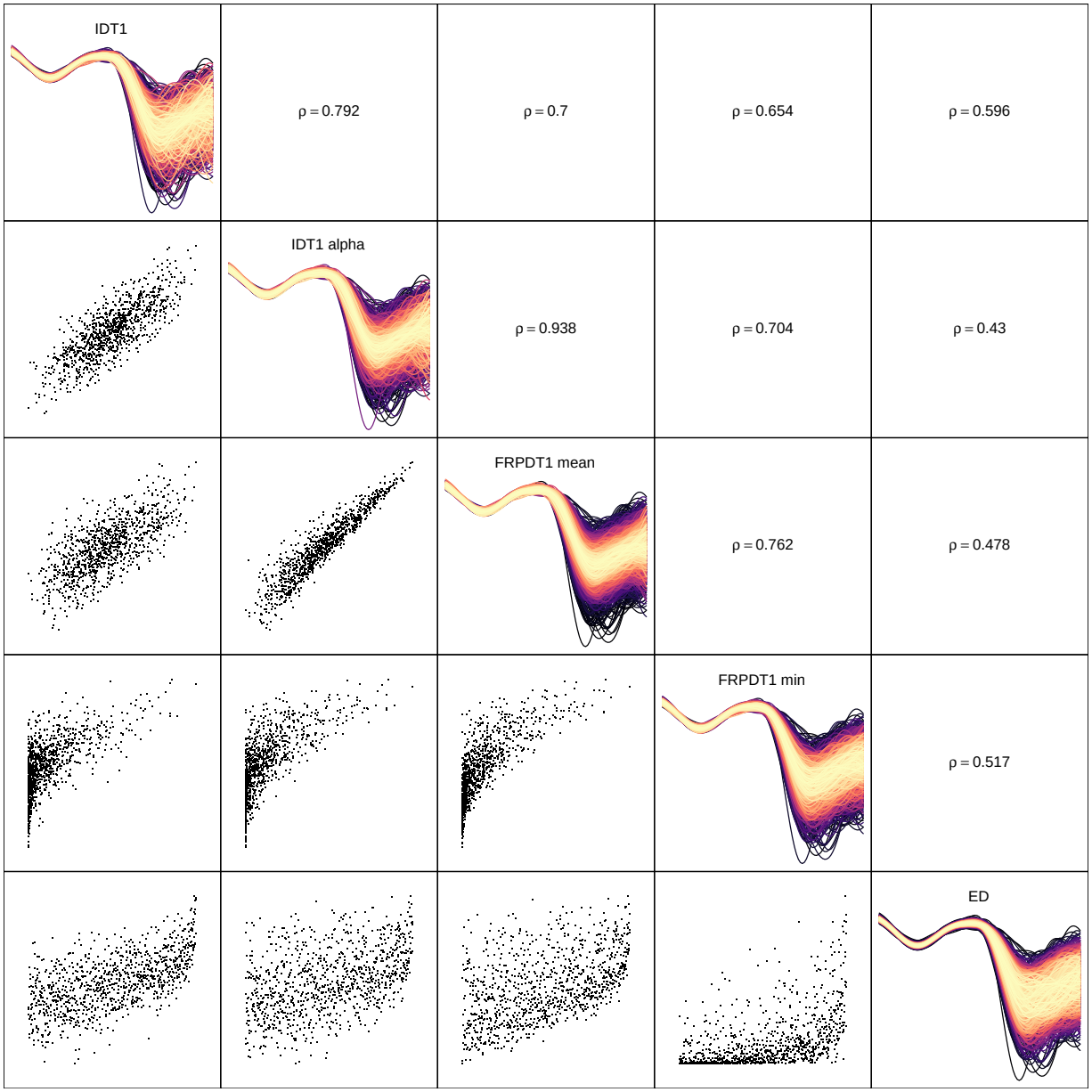


Figure 4.8: Matrix of the DD-plots for all pairs of D_{IDT1}, $D_{IDT1,\alpha}$, $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and D_{ED}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Gaussian process with varying amplitude.

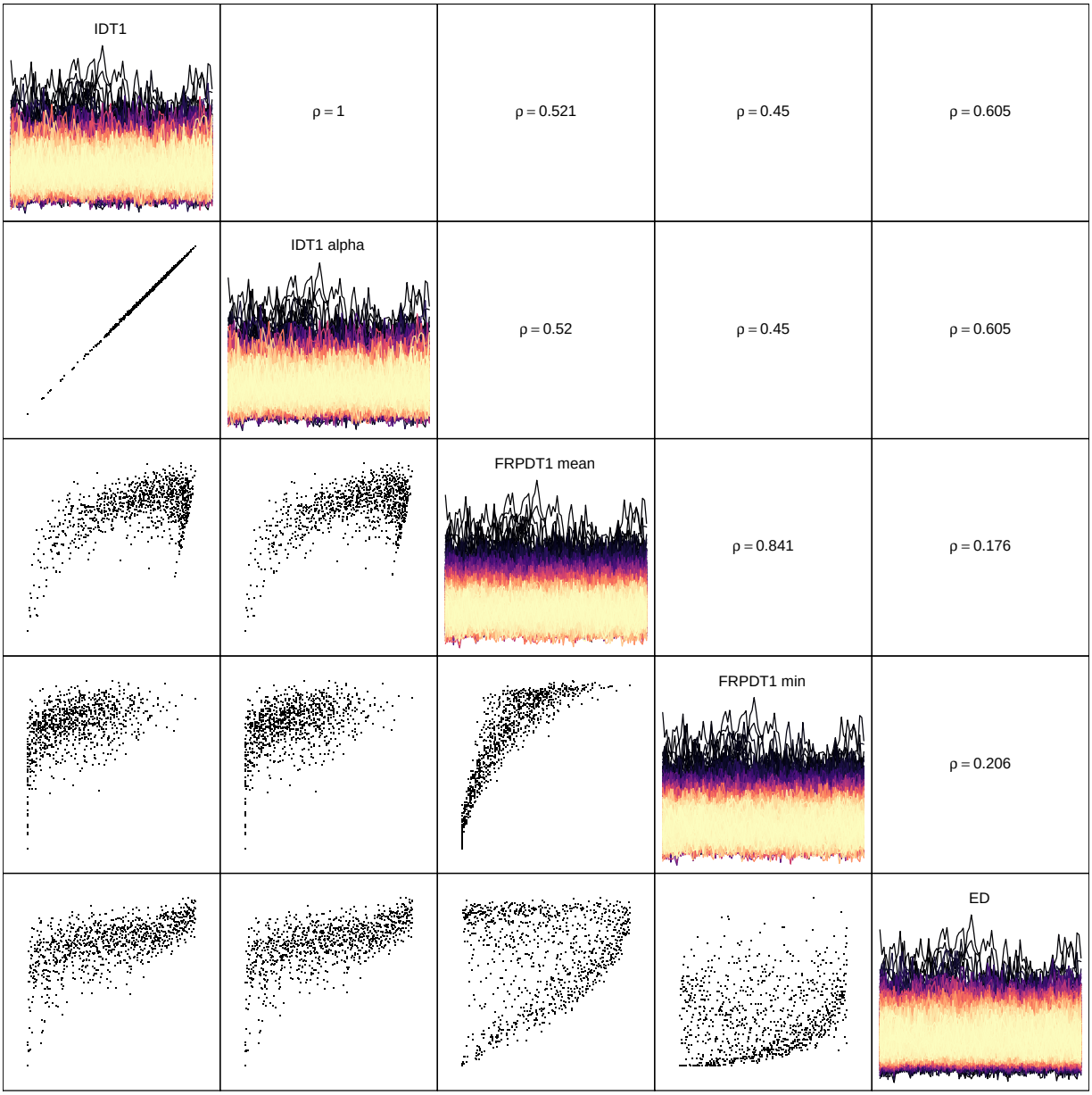


Figure 4.9: Matrix of the DD-plots for all pairs of D_{IDT1}, $D_{IDT1,\alpha}$, $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and D_{ED}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: Brown-Resnick (logarithms).

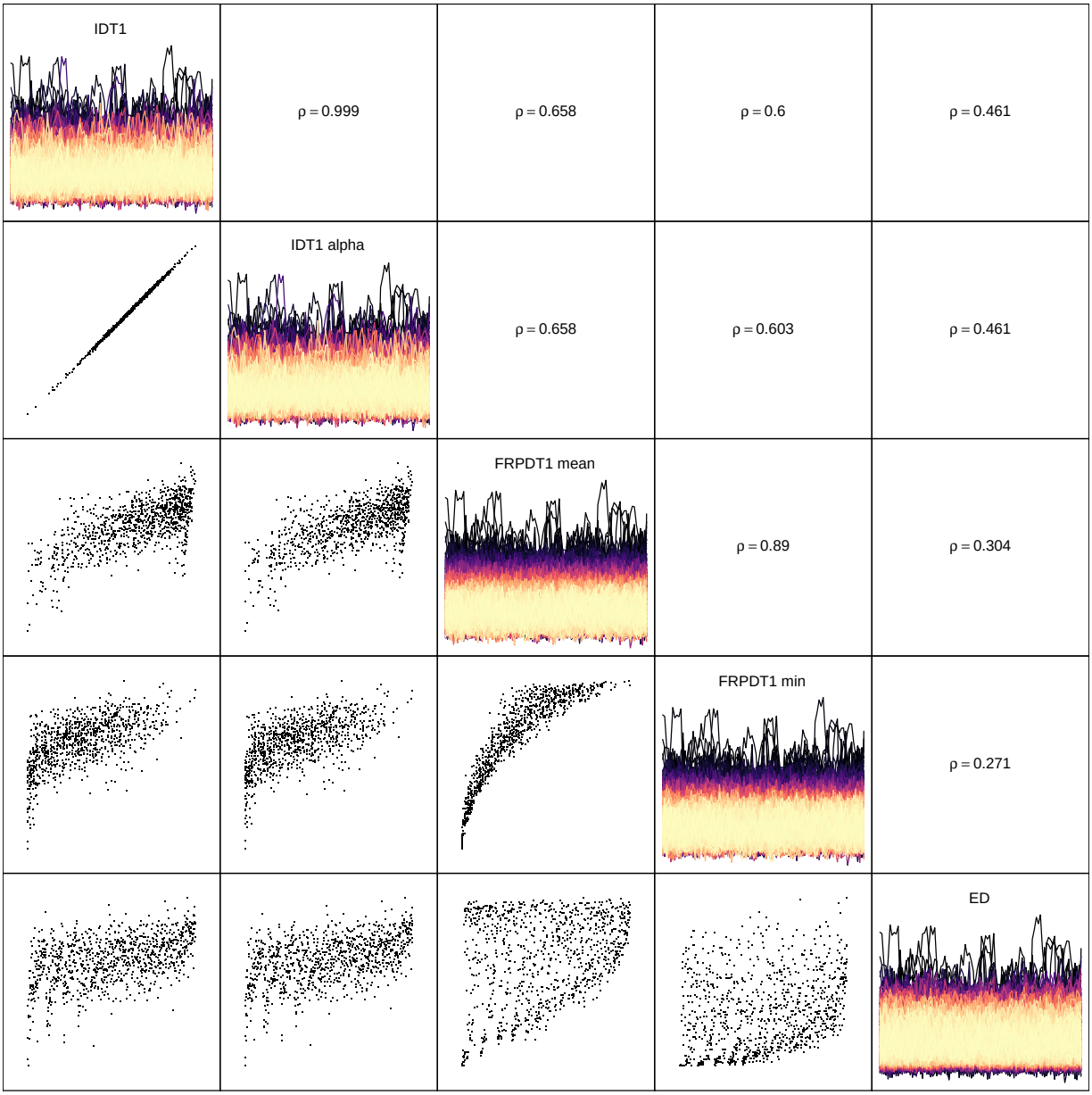


Figure 4.10: Matrix of the DD-plots for all pairs of D_{IDT1}, $D_{IDT1,\alpha}$, $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and D_{ED}, with corresponding correlation coefficients. On the diagonals the data is plotted and colored by depth: dark for extreme, light for deep. Data: extremal-t (logarithms).

It is interesting to observe that the data coloring and correlations patterns in the DD-plots all suggest very different mechanisms between the considered depth functions. In particular, for the functional random projection depth with minimum applied to Gaussian processes, the corresponding DD-plots show signs of possible degeneracy issues as the depth function seems to assign a depth close to zero to many data points.

The highest similarity is observed between the integrated depth with weight function and the functional random projection depth with mean when applied to Gaussian processes. More importantly, the similarity of the two depth functions on the Gaussian processes with varying amplitude adds further evidence to our conjecture that the functional random projection depth with mean satisfies Property P5. With regard to this property, this depth function even seems to perform better than the integrated depth with weight function. This is indicated by the narrower amplitude of the light region in the high amplitude part in its depth visualization. On the other hand, our conjecture about Property P5 with respect to the functional random projection depth with minimum is weakened by the observations of its depth visualization and its correlation to the integrated depth with weight function. Given the nature of the extremal depth, it is not surprising to see dark "envelopes" for this depth function over the range U on all data sets. On the Gaussian data with varying amplitude, these dark envelopes confirm our previous conjectures that the extremal depth does not satisfy Property P5.

We further notice, that when applied to max-stable processes, both versions of the functional random projection depth highly correlate. When comparing these depth functions to the integrated depth and the extremal depth, the DD-plots show completely different data orderings.

Next, we assess the performance of the integrated depth with weight function, both functional random projection depths and the extremal depth when applied to max-stable distributions to identify outliers. In light of possible extreme value theory applications, we are only considering the depth values themselves and thus are not able to incorporate the additional tool provided by the total variation decomposition of the integrated depth here.

For this study, we simulate five independent data sets from Brown-Resnick and extremal-t processes each, using the same parameters as in our previous simulations. For each data set, we then calculate the five most extreme functions for each depth, i.e. the five functions with the smallest respective depth values, and plot these next to each other. Additionally, as reference, we also plot the complete data set. The results are shown in the following figures.

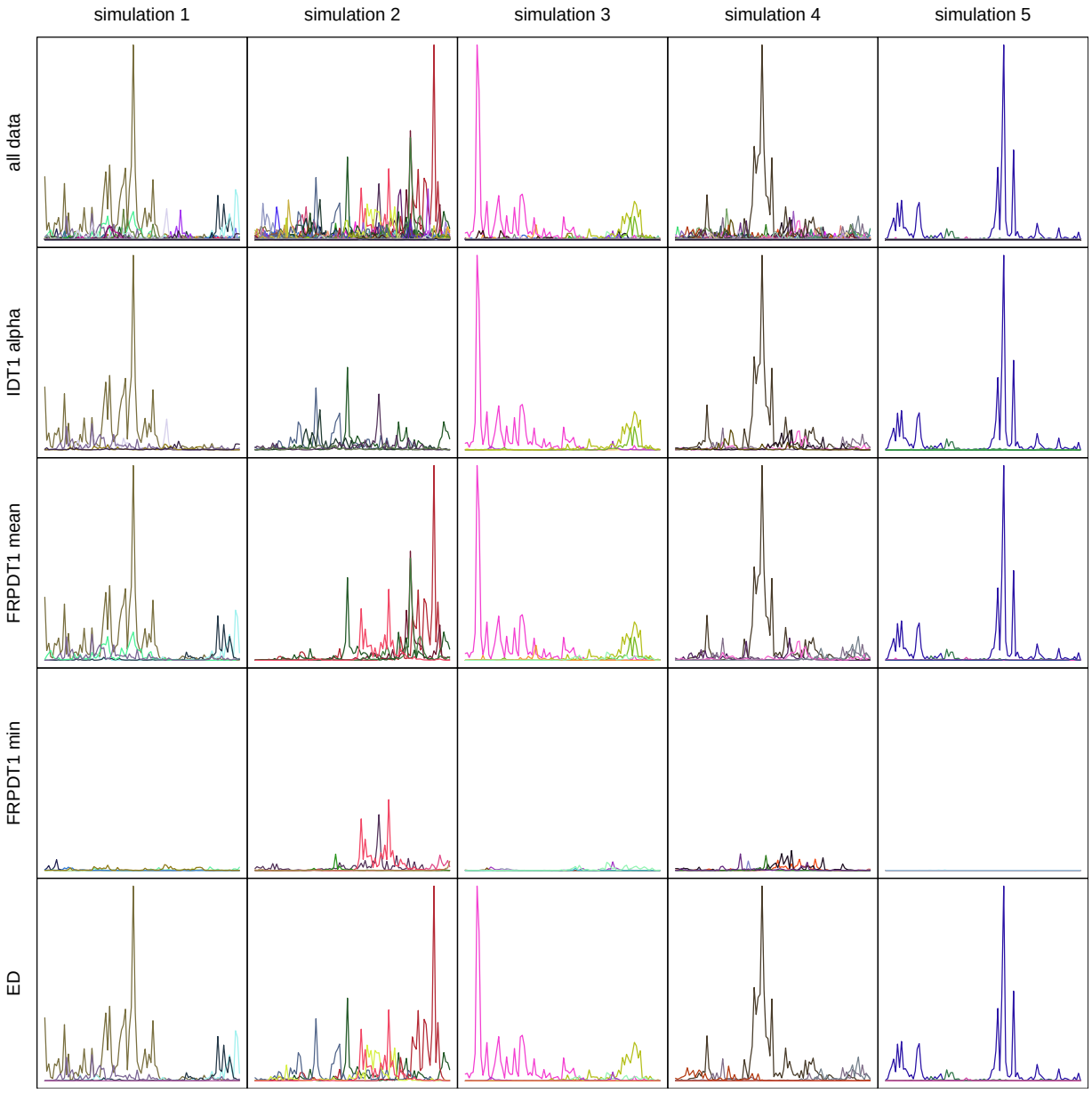


Figure 4.11: Plots of five $n = 1\,000$ Brown-Resnick data simulations (top row) and, respectively, plots of the five most extreme observations based on $D_{IDT1,\alpha}$, $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and D_{ED} calculations (each following row relates to one depth function). Colors are chosen randomly, though consistency is kept across plots related to the same data.

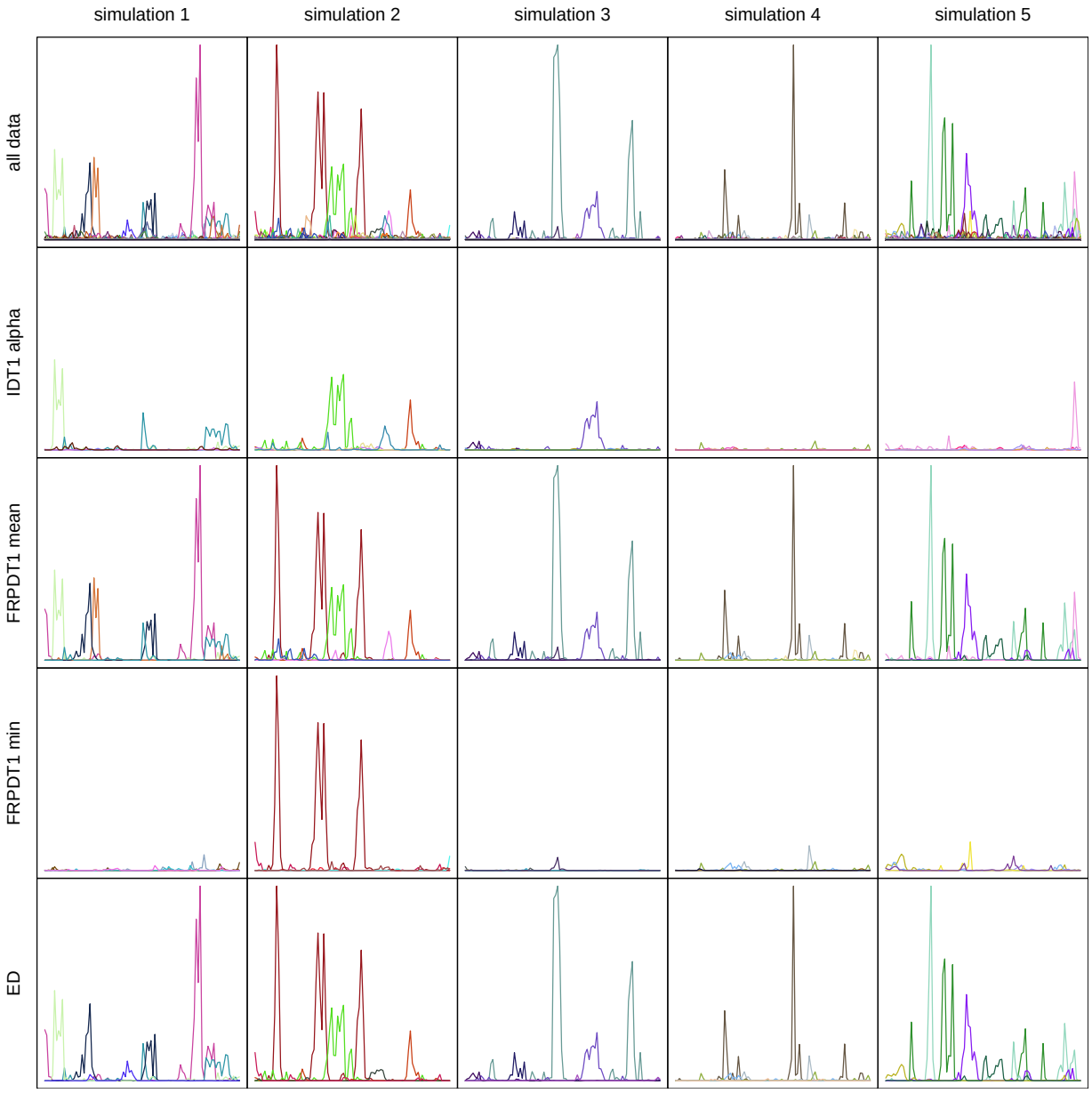


Figure 4.12: Plots of five $n = 1\,000$ extremal-t data simulations (top row) and, respectively, plots of the five most extreme observations based on $D_{IDT1,\alpha}$, $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and D_{ED} calculations (each following row relates to one depth function). Colors are chosen randomly, though consistency is kept across plots related to the same data.

The functional random projection depth with mean and the extremal depth seem to perform very well at identifying the five "most extreme" processes, as is implied by the similarity in the plots of the whole sample versus the corresponding plots with the five smallest depth values. The integrated depth, however, has difficulties identifying processes with important spikes, as we in particular observe on the extremal-t data. This is caused by the integration part of this depth function. Although a spike leads to a small marginal depth value on the specific interval, these low depth values then get "averaged out" over the interval U by the integral. When it comes to the functional random projection depth with minimum, its performs even worse, as for neither the Brown-Resnick, nor the extremal-t data sets, "extremes" are consistently identified. The reason behind it is that the minimum function in the depth construction, similarly as in the random Tukey depth, assigns the same smallest possible depth of $1/n$ to many curves. Since more than 5 processes have this depth value, no differentiation is possible anymore and the algorithm chooses 5 processes randomly amongst them. As noted previously, this phenomenon is also observed in Figure 4.7 and Figure 4.7, where one observes many depth values being close to zero.

Now, in the following chapter, we introduce some primary results of extreme value theory (EVT), found for example in de Haan and Ferreira (2006), and use them to assess possible modifications of existing depth functions and eventually potentially some of the mentioned degeneracy issues.

5 Modification of Depth Functions

5.1 Extreme Value Theory

5.1.1 Univariate Case

Consider $X_1, \ldots, X_n$, $n \in \mathbb{N}$, to be independent and identically distributed univariate random variables from a distribution P with mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let F denote the distribution function of P. The central limit theorem states that the average $\bar{X}_n := \sum_{i=1}^n X_i/n$ converges, with specific scaling, in distribution to the standard normal distribution

$$\frac{\bar{X}_n - d_n}{c_n} \xrightarrow{d} \mathcal{N}(0,1) \text{ as } n \to \infty,$$

where $d_n = \mu$, $c_n = \sqrt{\sigma^2/n}$ and $\xrightarrow{d}$ denotes convergence in distribution.

In extreme value theory, one is interested in similar convergence results for the sample maximum or minimum. For this, consider again independent and identically distributed univariate random variables $X_1, \ldots, X_n$, $n \in \mathbb{N}$ with distribution function. (Note that we make no assumptions on the existence of moments now.) Let $x^* := \sup\{x : F(x) < 1\}$ denote the right endpoint of the distribution F. Define the sample maximum by $M_n := \max(X_1, \ldots, X_n)$ and its corresponding distribution by

$$P(M_n \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = F^n(x),$$

where the last step holds because the random variables are independent and identically distributed. As a result, for $n \to \infty$, the distribution of the sample maximum converges in probability to a degenerate distribution function

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 0 & \text{for } x < x^* \\ 1 & \text{for } x \geq x^*, \end{cases}$$

if $x^* < \infty$. If $x^* = \infty$ then $\lim F^n(x) \equiv 0$ for $n \to \infty$. To alleviate this issue, in a similar approach to the central limit theorem, we seek to find sequences of constants $a_n > 0$, and b_n for $n = 1, 2, \ldots$, such that $(M_n - b_n)/a_n$ has a non-degenerate limit distribution as $n \to \infty$. Expressed differently, such that

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \tag{5.1}$$

holds for every continuity point x of G, where G is a non-degenerate distribution function. If (5.1) holds, we say that F is in the *domain of attraction* of G ($DoA(G)$) and G is called an *extreme value*

distribution. Any extreme value distribution G can be written as

$$G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-\frac{1}{\gamma}}\right\} \quad \text{for } 1+\gamma x > 0 \text{ and } \gamma \in \mathbb{R},$$

where the index γ is called the *extreme value index*. For this part of the thesis, we are mainly interested in the class of heavy tailed distributions, that is distributions with $\gamma > 0$. Examples of such distributions are the Cauchy or the Fréchet distribution. The latter is defined by

$$\Xi_\alpha(x) := \exp(-x^{-\alpha}) \quad \text{for } \alpha > 0 \text{ and } x > 0. \tag{5.2}$$

Next, we consider the right tail of the distribution function F defined by

$$p_r(x) := 1 - F(x) = P(X > x). \tag{5.3}$$

Since for the general nonparametric approach, the distribution function F is estimated by the empirical distribution function F_n, p_r is estimated zero outside of the sample range, i.e. in particular for large values of x. Thus, an alternative method for the estimation of the tail is required. Replacing $n \in \mathbb{N}$ by $t \in \mathbb{R}$ and using that $-\log(1-x) \sim x$ for x approaching zero from above, we can write (5.1) as

$$\lim_{t\to\infty} t\left(1 - F\big(a(t)y + b(t)\big)\right) = -\log G_\gamma(y) = (1+\gamma y)^{-1/\gamma},\ 1+\gamma y > 0. \tag{5.4}$$

Setting $t = n/\kappa$, $\kappa \in \mathbb{N}$ and $x = a(t)y + b(t)$ in (5.4), we obtain the following approximation for large n/κ and large x

$$p_r(x) \approx \frac{\kappa}{n}\left(1 + \gamma\frac{x - b(n/\kappa)}{a(n/\kappa)}\right)^{-1/\gamma}. \tag{5.5}$$

Now, to obtain an estimator for $p_r(x)$, we first require an estimator for γ.

Definition 5.1. The *Hill estimator* (Hill, 1975) of the extreme value index γ, for $\gamma > 0$ is defined by

$$\hat{\gamma}(X_1, \ldots, X_n) := \frac{1}{\kappa}\sum_{i=0}^{\kappa-1} \log(X_{n-i:n}) - \log(X_{n-\kappa:n}), \tag{5.6}$$

where $X_{i:n}$ denotes the i-th order statistic of $X_1, \ldots, X_n$.

Further, in de Haan and Ferreira (2006), it is shown that when $\gamma > 0$, $\hat{a}(n/\kappa) := \hat{\gamma}_r X_{n-\kappa:n}$ and $\hat{b}(n/\kappa) := X_{n-\kappa:n}$ can be used as estimator for, respectively, $a(n/\kappa)$ and $b(n/\kappa)$, where $\hat{\gamma}_r := \hat{\gamma}(X_1, \ldots, X_n)$. Plugging these estimators into (5.5), we obtain the following estimator for the right tail probability

$$p_{r,n}(x) := \frac{\kappa}{n}\Big(\frac{x}{X_{n-\kappa:n}}\Big)_+^{-1/\hat{\gamma}_r},$$

where $c^+ := \max\{0, c\}$.

Since a minimum can be reformulated into a maximum by

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n),$$

we can define the left-tail probability of the distribution F, $p_l(x)$, similarly as $p_r(x)$ and estimate it by

$$p_{l,n}(x) := \frac{\kappa}{n}\Big(\frac{x}{-X_{\kappa:n}}\Big)_+^{-1/\hat{\gamma}_l},$$

where $\hat{\gamma}_l := \hat{\gamma}(-X_1, \ldots, -X_n)$.

With this, an extreme value theory corrected estimator for the cumulative distribution function F can be defined (e.g. de Haan and Ferreira (2006)).

Definition 5.2. Let $X_1, \ldots, X_n$ be independent univariate random variables with a non-degenerate distribution function F and let $\gamma > 0$ in both tails. An extreme value theory corrected estimator of the distribution function F is defined by

$$F_{n,EVT}(x) := \begin{cases} 1 - \frac{\kappa}{n}\Big(\frac{x}{-X_{\kappa:n}}\Big)_+^{-1/\hat{\gamma}_l} & \text{for } x \leq X_{\kappa:n} \\ F_n(x) & \text{for } X_{\kappa:n} < x < X_{n-\kappa:n} \\ 1 - \frac{\kappa}{n}\Big(\frac{x}{X_{n-\kappa:n}}\Big)_+^{-1/\hat{\gamma}_r} & \text{for } x \geq X_{n-\kappa:n}, \end{cases} \tag{5.7}$$

where $F_n(x)$ is the empirical distribution defined in (3.4).

Remark 5.3. Sometimes, one is only interested in modifying the right tail of the distribution. In this case, we denote the modified estimator for the distribution function F by $F_{r,n,EVT}$. This estimator differs from $F_{n,EVT}$ defined in (5.7) for $x \leq X_{\kappa:n}$, where $F_{r,n,EVT}(x) = F_n(x)$ holds.

Another important concept in extreme value theory, in particular for later generalizations to the multivariate case, is the regularly varying condition on $\mathbb{R}$.

Definition 5.4. A measurable function $f\colon \mathbb{R}_+ \longmapsto \mathbb{R}_+$ is regularly varying at infinity with index α, written $f \in RV_\alpha$, if for $x > 0$

$$\lim_{t\to\infty} \frac{f(tx)}{f(t)} = x^\alpha.$$

For a random variable, the regularly varying condition is linked to the domain attraction condition in the following way.

Theorem 5.5. (Gnedenko, 1943)

With the notation introduced above, it holds that $F \in DoA(G_\gamma)$ for $\gamma > 0$ if and only if $1 - F \in RV_{-\alpha}$ with $\alpha = 1/\gamma$.

5.1.2 Multivariate Case

We now consider the multivariate case, where analogously to the univariate case, a multivariate domain of attraction condition and a multivariate regular varying condition can be defined, see for example

de Haan and Ferreira (2006) and Resnick (1987).

Definition 5.6. Let $\boldsymbol{X_1}, \dots, \boldsymbol{X_n}$ be independent K-variate random vectors with distribution function F, $n \in \mathbb{N}$, $K \in \mathbb{N}$ and write $\boldsymbol{X_j} = (X_j^{(1)}, \dots, X_j^{(K)})$, $j = 1, \dots, n$. Further, let the marginal distributions of $F(\boldsymbol{x})$, $\boldsymbol{x} = (x^{(1)}, \dots, x^{(K)}) \in \mathbb{R}^K$, be $F_1, \dots, F_K$ so that $F_1(x^{(1)}) = F(x^{(1)}, \infty, \dots, \infty)$ and so on. Assume that there exist normalizing constants $a_n^{(i)} > 0$, $b_n^{(i)} \in \mathbb{R}$, $i = 1, \dots, K$, $n \geq 1$ such that as $n \to \infty$

$$\begin{aligned} &P\big[(M_n^{(i)} - b_n^{(i)})/a_n^{(i)} \leq x^{(i)}, i = 1, \dots, K\big] \\ &\qquad = F^n(a_n^{(1)}x^{(1)} + b_n^{(1)}, \dots, a_n^{(K)}x^{(K)} + b_n^{(K)}) \xrightarrow{d} G(\boldsymbol{x}), \end{aligned} \tag{5.8}$$

where $M_n^{(i)} = \max\{X_1^{(i)}, \dots, X_n^{(i)}\}$ is the i-th marginal sample maximum. Here, the class of limit distributions G are called *multivariate extreme value distributions* and any F giving rise to (5.8) is said to be in the *domain of attraction* of G, written $F \in DoA(G)$.

Definition 5.7. Let $\boldsymbol{X}$ be a K-dimensional random vector with distribution P, $K \in \mathbb{N}$. Then its distribution is said to be *multivariate regularly varying with index* $\alpha > 0$, if there exists a random vector $\boldsymbol{\Theta}$ with values in $\mathbb{S}^{K-1}$ with respect to the Euclidiean norm $\|\cdot\|$, such that for all $t > 0$,

$$\frac{P(\|\boldsymbol{X}\| > tx, \boldsymbol{X}/\|\boldsymbol{X}\| \in \cdot\,)}{P(\|\boldsymbol{X}\| > t)} \longrightarrow t^{-\alpha} P(\Theta \in \cdot\,), \quad \text{as } x \to \infty.$$

If all marginals of P are the same, it is shown in Resnick (1987), Proposition 5.15 that also the multivariate regular varying condition is equivalent to the multivariate domain of attraction condition. Further, in Basrak et al. (2002), it is shown that a K-variate random vector $\boldsymbol{X}$ is multivariate regularly varying if and only if every linear combination $v'X$, $v \in \mathbb{R}^K$, is regularly varying on $\mathbb{R}$, where v' denotes the transposed version of v. This result enables us to apply the extreme value theory corrected empirical cumulative distribution function defined in (5.7) to multivariate data.

5.1.3 Functional Case

For the functional case, the domain of attraction condition is defined as follows.

Definition 5.8. Consider $X_1, \dots, X_n$ to be independent replications of a stochastic process with distribution function F with paths on $\mathcal{C}(U), U \subseteq \mathbb{R}$, $n \in \mathbb{N}$ and continuous marginals F_t. Assume that there exist normalizing functions $a_n(t)$ positive and $b_n(t)$ real such that as $n \to \infty$

$$\left\{ \max_{i=1,\dots,n} \frac{X_i(t) - b_n(t)}{a_n(t)} : t \in U \right\} \xrightarrow{d} \{G(t) : t \in U\}. \tag{5.9}$$

Any F giving rise to (5.9) is said to be in the *domain of attraction* of G, written $F \in DoA(G)$.

Now come to the definition of a max-stable process.

Definition 5.9. A stochastic process G on $\mathcal{C}(U)$ with non-degenerate marginals is called *max-stable* if there exist continuous functions $a_n(t)$ positive and $b_n(t)$ real such that if $G_1, \ldots, G_n$ are independent and identical copies of G, then

$$\max_{i=1,\ldots,n} \left\{ \frac{G_i - b_n}{a_n} \right\} \stackrel{d}{=} G \tag{5.10}$$

for $n = 1, 2, \ldots$

One primary results of extreme value theory is that this class of max-stable processes is shown to coincide with the class of limit processes in (5.9). Finally, we define a special case of max-stable processes.

Definition 5.10. A stochastic process on $\mathcal{C}^+(U)$ with non-degenerate marginals is called *simple max-stable* if (5.10) holds and $P(G(t) \leq 1) = \exp(-1)$ for all $t \in U$, i.e. G has standard Fréchet marginals, where $\mathcal{C}^+(U)$ denotes the space of continuous and positive functions on U.

5.2 Modified Random Tukey Depth

In Einmahl et al. (2015), the introduced extreme value theory concepts are applied to modify the Tukey depth and the random Tukey depth for multivariate data defined, respectively, in (3.5) and (3.6). For each projection onto $v_j \in \mathbb{R}^K$, $j = 1, \ldots, k$, instead of using the marginal empirical distribution function $F_{v_j,n}$ corresponding to the distribution $P_{v_j,n}$ to estimate the marginal distributions, the modified extreme value theory version from (5.7) is used. The result is the following definition.

Definition 5.11. (Einmahl et al., 2015)

Let ν be the uniform distribution on the unit sphere on $\mathbb{R}^K$ and $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$ be independent and identically distributed multivariate regularly varying random vectors with tail index $\alpha > 0$. *The extreme value theory corrected empirical random Tukey depth* of $\boldsymbol{x} \in \mathbb{R}^K$ with respect to P_n and based on k realizations of independent random vectors $\boldsymbol{v_1}, \ldots, \boldsymbol{v_k}$ chosen with distribution ν is

$$D_{RT,EVT,k,\nu}(\boldsymbol{x}, P_n, \hat{\alpha}) := \min_{j=1,\ldots,k} \left\{ F_{\boldsymbol{v_j},n,EVT}(\boldsymbol{v_j'x}, \hat{\alpha}), 1 - F_{\boldsymbol{v_j},n,EVT}(\boldsymbol{v_j'x}, \hat{\alpha}) \right\}, \tag{5.11}$$

where, for any $\boldsymbol{v} \in \mathbb{R}^K$ and $W_i = \boldsymbol{v'X_i}$ we have

$$F_{\boldsymbol{v},n,EVT}(w, \hat{\alpha}) := \begin{cases} 1 - \frac{\kappa}{n} \left(\frac{w}{-W_{\kappa:n}} \right)_+^{-\hat{\alpha}} & \text{for } w \leq W_{\kappa:n} \\ F_n(w) & \text{for } W_{\kappa:n} < w < W_{n-\kappa:n} \\ 1 - \frac{\kappa}{n} \left(\frac{w}{W_{n-\kappa:n}} \right)_+^{-\hat{\alpha}} & \text{for } w \geq W_{n-\kappa:n}. \end{cases} \tag{5.12}$$

Further, the tail index $\alpha = 1/\gamma$ is estimated by

$$\hat{\alpha} = 1/\hat{\gamma}(\|\boldsymbol{X_1}\|, \ldots, \|\boldsymbol{X_n}\|) \tag{5.13}$$

where $\|\cdot\|$ is the Euclidian norm on $\mathbb{R}^K$.

Remark 5.12. The estimator $F_{\boldsymbol{v_j},n,EVT}$ from (5.12) is the estimator $F_{n,EVT}$ from (5.7) applied to the marginal distribution $(\boldsymbol{v_j'x_1},\dots,\boldsymbol{v_j'x_n})$ with $\hat{\alpha} = 1/\gamma_r = 1/\gamma_l$ for each $\boldsymbol{v_j} \in \{\boldsymbol{v_1},\dots,\boldsymbol{v_k}\}$.

In the next section, we study the introduced depth function for selected simulations.

5.3 Simulation Study for Multivariate Data

To analyze the extreme value theory corrected random Tukey depth, we compare it to the random Tukey depth and the random projection depth for three distribution functions presented in He and Einmahl (2017).

1. the bivariate Cauchy distribution ($\gamma = 1$) with density

$$f(x,y) := \frac{1}{2\pi(1+x^2+y^2)^{3/2}}, \qquad (x,y) \in \mathbb{R}^2;$$

2. a bivariate "clover" distribution ($\gamma = 1/3$) with density

$$f(x,y) := \frac{3\left[9(x^2+y^2)^2 - 32x^2y^2\right]}{10\pi\left[1+(x^2+y^2)^3\right]^{3/2}}, \qquad (x,y) \in \mathbb{R}^2;$$

3. a bivariate elliptical distribution ($\gamma = 1/3$) with density

$$f(x,y) := \frac{3(x^2/4+y^2)^2}{4\pi\left[1+(x^2/4+y^2)^3\right]^{3/2}}, \qquad (x,y) \in \mathbb{R}^2.$$

For each function, we simulate samples of size $n = 1\,000$. The first step requires to find a good value κ for the Hill estimator of γ implied by (5.13). For this purpose, we consider *Hill plot* (Hill, 1975): Here, the Hill estimator $\hat{\gamma}$ is plotted in dependence of κ. One then examines its variability and the first interval over which the variability is low provides a good reference for a good value κ. Figure 5.1 shows nine Hill plots for nine Monte Carlo simulations.

Once an interval with low variability is found, we repeat the estimation process for $m = 100$ Monte Carlo simulations, and take the average over all Hill estimators. We obtain $\hat{\gamma} = 1.008$ as the final value. For further calculations, a value of $\kappa = 175$ is used.

We apply this process to the other two distribution functions to obtain $\hat{\gamma} = 0.3414$ and $\kappa = 225$ for the clover distribution and $\hat{\gamma} = 0.3447$ and $\kappa = 225$ for the elliptical distribution. The corresponding plots are found in Figure 8.3 and Figure 8.4 in the Appendix.

Now, for each distribution function we apply the extreme value theory corrected random Tukey depth, the random Tukey depth and the random projection depth to the simulated data and plot the obtained depth values pairwise against each other in DD-plots. The results are shown in Figure 5.2, Figure 5.3 and Figure 5.4.

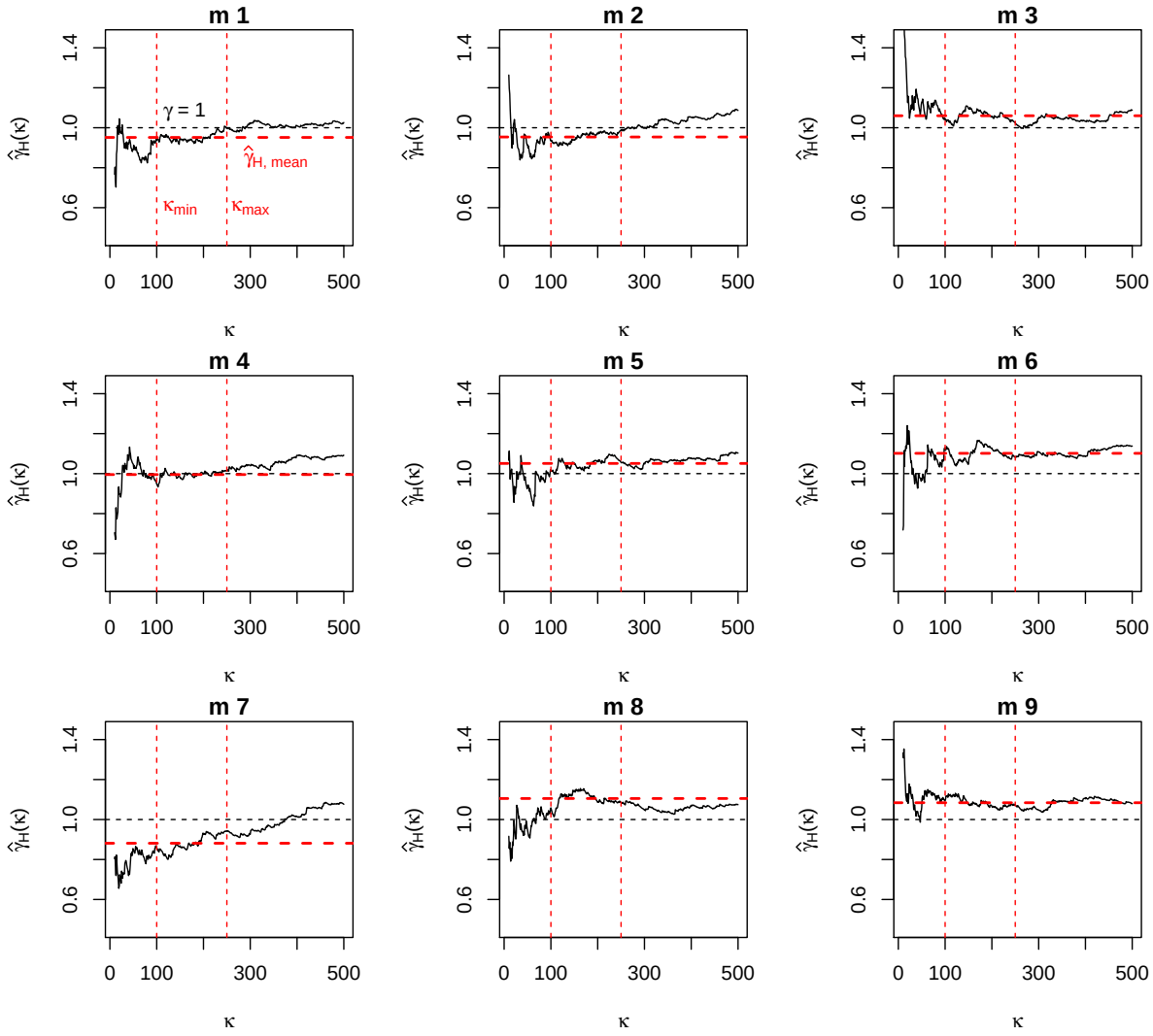


Figure 5.1: Hill plots for nine Monte Carlo simulations, each with $n = 1\,000$ realizations of bivariate Cauchy data. The interval, where variability is low over all plots is estimated to be $[100, 250]$. On each plot, the horizontal dashed red line indicates the average Hill estimator over this interval and the horizontal dashed black line indicates the true value of gamma, here $\gamma = 1$

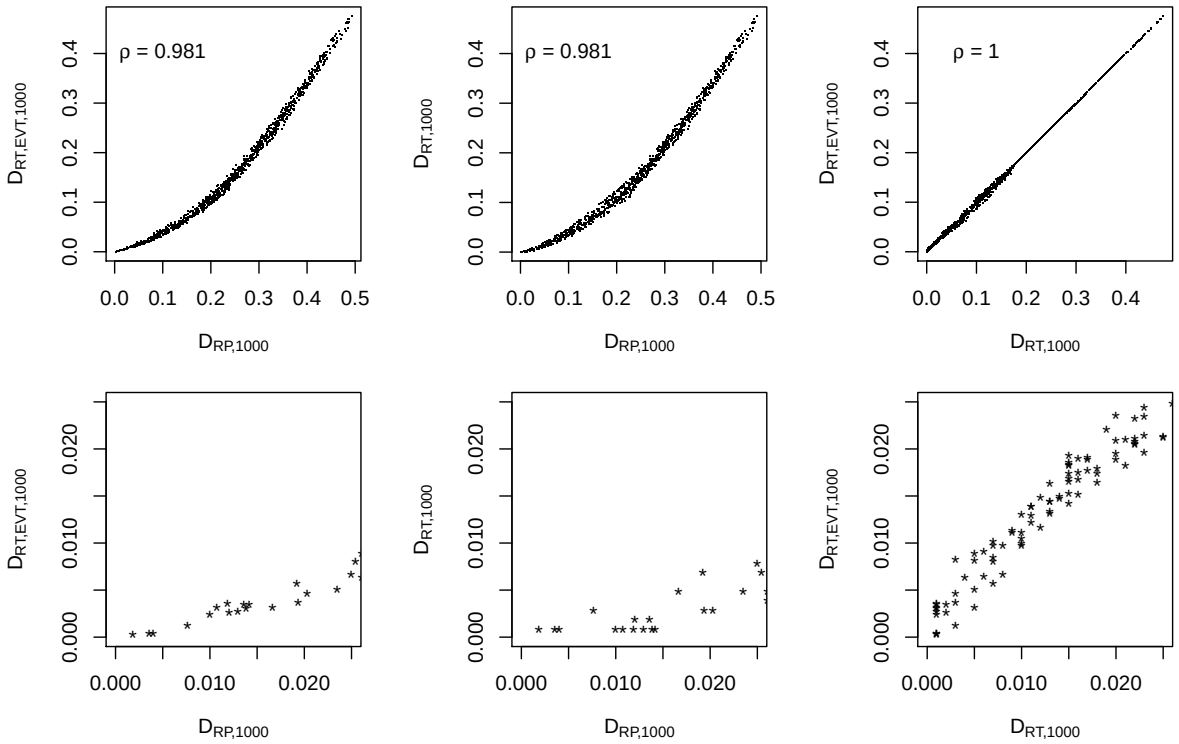


Figure 5.2: Pairwise DD-plots between the extreme value theory corrected random Tukey depth ($D_{RT,EVT}$), the random Tukey depth (D_{RT}) and the random projection depth (D_{RP}) for $n = 1\,000$ realizations of bivariate Cauchy random vectors and $k = 1\,000$ random projections. The top three plots show full DD-plots, while the lower plots show a corresponding version that is zoomed to the lower left corner, i.e. where depth values are usually small for both depth functions.

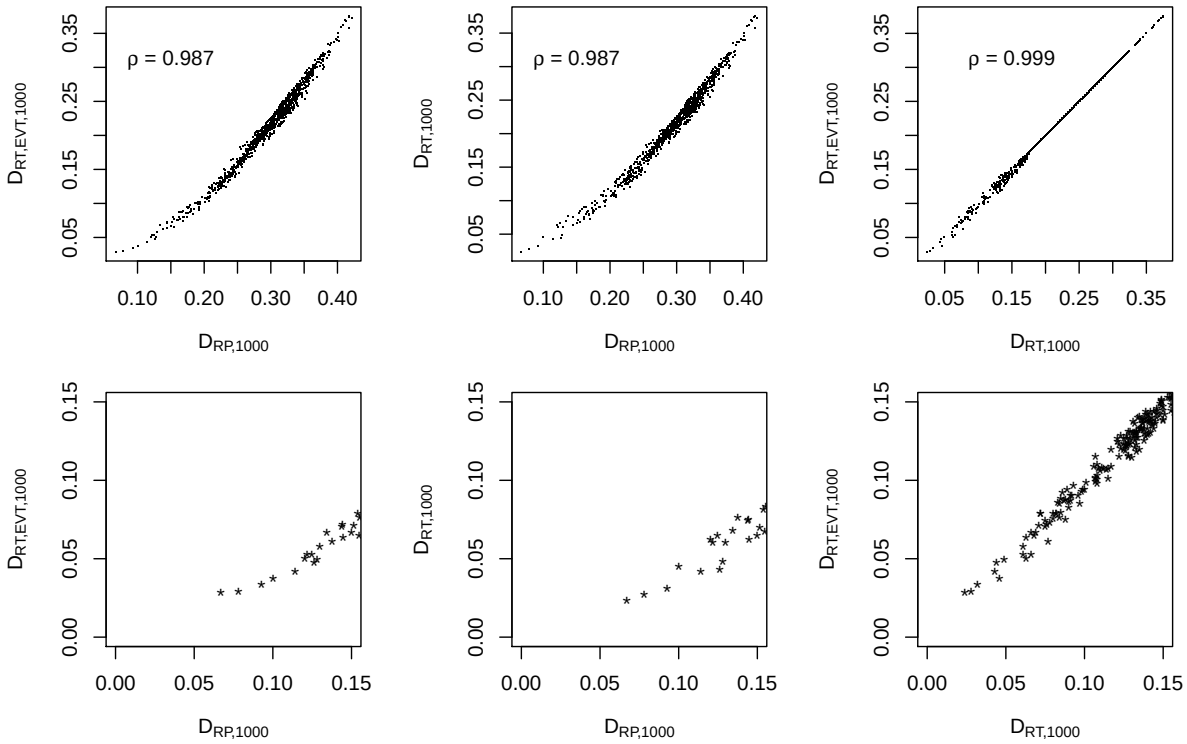


Figure 5.3: Pairwise DD-plots between the extreme value theory corrected random Tukey depth ($D_{RT,EVT}$), the random Tukey depth (D_{RT}) and the random projection depth (D_{RP}) for $n = 1\,000$ realizations from a bivariate clover distribution and $k = 1\,000$ random projections. The top three plots show full DD-plots, while the lower plots show a corresponding version that is zoomed to the lower left corner, i.e. where depth values are usually small for both depth functions.

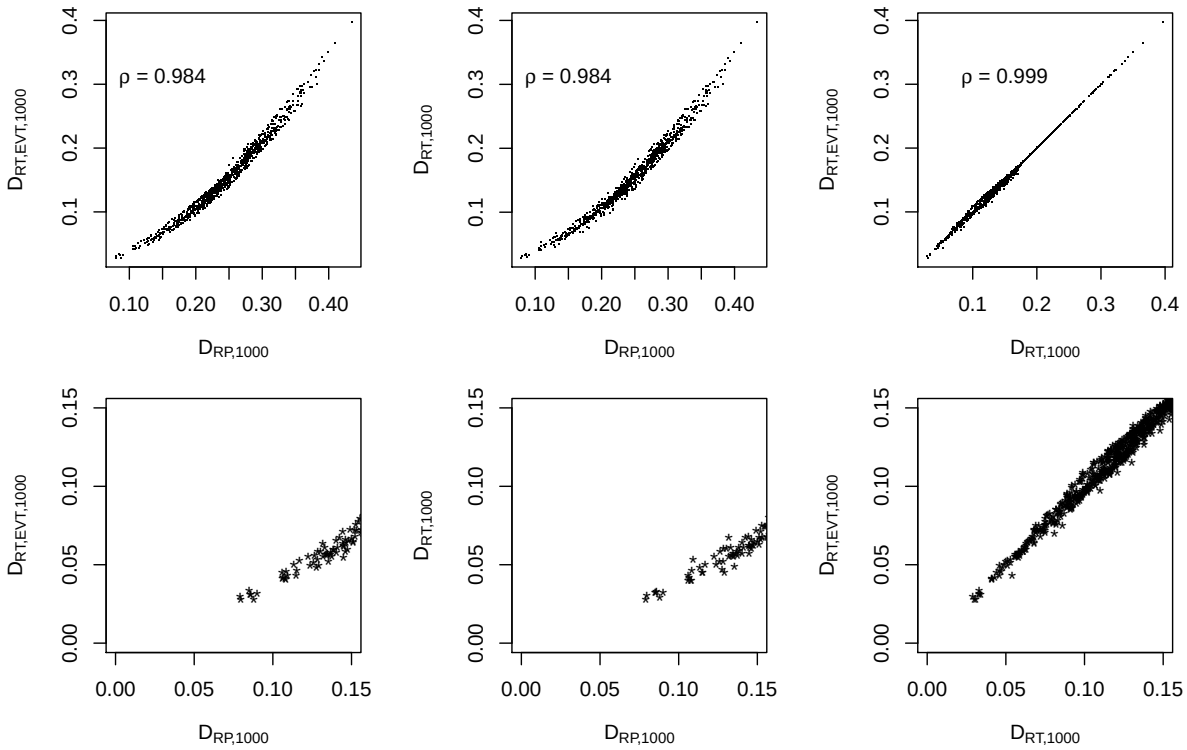


Figure 5.4: Pairwise DD-plots between the extreme value theory corrected random Tukey depth ($D_{RT,EVT}$), the random Tukey depth (D_{RT}) and the random projection depth (D_{RP}) for $n = 1\,000$ realizations from a bivariate elliptical distribution and $k = 1\,000$ random projections. The top three plots show full DD-plots, while the lower plots show a corresponding version that is zoomed to the lower left corner, i.e. where depth values are usually small for both depth functions.

From these figures, we notice a high similarity between the extreme value theory corrected random Tukey depth and the random projection depth, even in the tails (i.e. the lower left part of the corresponding DD-plots) for all three data simulations. While we observe some sign of degeneracy for the Cauchy data when using the classical random Tukey depth, this is not the case when using the other two depth functions. Since the clover distribution and the elliptical distribution have a smaller γ ($\gamma = 1/3$ for both) than the Cauchy distribution ($\gamma = 1$), they feature lighter tails and in this setting it seems that correcting the random Tukey depth does not affect improve the depth function much.

We now consider the depth functions for functional data.

5.4 Modified Functional Random Projection Depth

To apply extreme value theory to functional depth functions, we consider the four depth functions remaining after our analysis in Section 4.3: the integrated depth with weight function, the functional random projection depth with mean and with minimum and the extremal depth. For each of these depth functions, the corresponding marginal depth function can be chosen as the Tukey depth and thus, a straightforward modification is possible. However, for the integrated depth, an extreme value theory modification might successfully modify the marginal distribution function F_t, $t \in U$, but since these modifications only happen in the tail, these changes get diluted with the integral over $t \in U$. Similarly, through the mean in the functional random projection depth, here too the modification leads to virtually identical final depth values. For the extremal depth, also due to the integral, an extreme value theory modification is unlikely to affect the final ranking. This only leaves the functional random projection depth with minimum to modify. Recall its definition.

Definition 3.13 Let D_M be any depth for univariate data and let η be a non-degenerate probability distribution on $\mathcal{F}$. The *functional random projection depth with minimum* of $x \in \mathcal{F}$ with respect to P based on k stochastic processes with distribution η is

$$D_{FRPD,k,\eta,\min}(x, P, D_M) := \min_{i=1,\dots,k} \Big\{ D_M(\langle v_i, x\rangle, P_{v_i}) \Big\}, \tag{5.14}$$

where $v_1, \dots, v_k$ are realizations of independent stochastic processes with distribution η and $\langle v_i, x\rangle$ is the $\mathcal{L}^2$-inner product between v_i and x, $i = 1, \dots, k$. Additionally, P_{v_i} denotes the marginal distribution of P on $\{\langle v_i, y\rangle : y \in \mathcal{F}\}$, $i = 1, \dots, k$.

For the empirical version, we replace P and P_{v_i} by, respectively, P_n and $P_{v_i,n}$ to obtain the following definition.

Definition 5.13. Let D_M be any depth for univariate data and let η be a non-degenerate probability distribution on $\mathcal{F}$. The *empirical functional random projection depth with minimum* of $x \in \mathcal{F}$ with respect to P based on k stochastic processes with distribution η is

$$D_{FRPD,k,\eta,\min}(x, P_n, D_M) := \min_{i=1,\dots,k} \Big\{ D_M(\langle v_i, x\rangle, P_{v_i,n}) \Big\}, \tag{5.15}$$

where $v_1, \dots, v_k$ are realizations of independent stochastic processes with distribution η and x, $i = 1, \dots, k$, $P_{v_i,n}$ is the marginal distribution on the random sample $\boldsymbol{Y}_{v_i,n}$ with

$$\boldsymbol{Y}_{v_i,n} := \Big\{ \langle v_i, X_1\rangle, \dots, \langle v_i, X_n\rangle \Big\}, \tag{5.16}$$

for $i = 1, \dots, k$.

If $X_1, \dots, X_n$ are independent and identical replications of a simple max-stable process and η is chosen as a realization of a Gaussian process, then we assume that the *absolute values* of their inner

products $\boldsymbol{Y}_{v_i,n}$ are regularly varying with index α for $i = 1, \dots, k$. The absolute values here are necessary, because while the simple max-stable processes have only positive values, the realization of the Gaussian process v_i can have positive and negative values and thus so can the values of the sample $\boldsymbol{Y}_{v_i,n}$. However, since the marginals of a simple max-stable distribution are Fréchet distributed, only their right tail is heavy-tailed. (Their left tail is finite at zero.) Depending on the sign of the majority of the values in the realizations v_i, the distribution of the sample $\boldsymbol{Y}_{v_i,n}$ can be heavy-tailed for only the right tail, only the left tail or both as is observed in the four examples plotted in Figure 5.5. While this simulation seems to confirm our regularly varying assumption, showing this assumption is beyond the scope of this thesis.

For the reasons detailed above, we define a new estimator for the distribution function F.

Definition 5.14. Let $X_1, \dots, X_n$ be independent random variables with distribution function F and let $|X_1|, \dots, |X_n|$ be regularly varying with index $\alpha > 0$, then we define the new extreme value theory corrected estimator for the distribution function F by

$$\tilde{F}_{n,EVT}(x, \hat{\alpha}) := \begin{cases} F_n(x) & \text{for } x < Z_{n-\kappa:n} \\ 1 - \frac{\kappa}{n}\left(\frac{x}{Z_{n-\kappa:n}}\right)_+^{-\hat{\alpha}} & \text{for } x \geq Z_{n-\kappa:n} \end{cases}, \tag{5.17}$$

where $Z_i = |X_i|$, $Z_{(i)}$ is the i-th order statistic of the sample $\{Z_1, \dots, Z_n\}$ and $\hat{\alpha}$ is an estimator for the tail index α.

One possible choice for an estimator of the index α is $\hat{\alpha} = 1/\hat{\gamma}(Z_1, \dots, Z_n)$, where $\hat{\gamma}(\cdot)$ is the Hill estimator defined in (5.6).

With this, we define an extreme value theory corrected empirical univariate Tukey depth.

Definition 5.15. The *extreme value theory corrected empirical univariate Tukey depth* of $x \in \mathbb{R}$ with respect to P_n is defined by

$$D_{T1,EVT}(x, P_n, \hat{\alpha}) := \min\left\{\tilde{F}_{n,EVT}(x, \hat{\alpha}), 1 - \tilde{F}_{n,EVT}(x, \hat{\alpha})\right\}, \tag{5.18}$$

where $\tilde{F}_{n,EVT}$ is the extreme value theory corrected estimator of the distribution function F defined in (5.17) and $\hat{\alpha}$ is an estimator for the tail index α.

With this, we can modify the functional random projection depth with minimum as follows.

Definition 5.16. Let $D_{T1,EVT}$ be the extreme value theory corrected empirical Tukey depth defined in (5.18) and let η be a Gaussian process on $\mathcal{L}^2(U)$. The *extreme value theory corrected empirical functional random projection depth with minimum* of $x \in \mathcal{F}$ with respect to P_n based on k stochastic processes with distribution η is

$$D_{FRPD,EVT,k,\eta,\min}(x, P_n, D_{T1,EVT}) := \min_{i=1,\dots,k}\left\{D_{T1,EVT}\left(\langle v_i, x\rangle, P_n, \hat{\alpha}_i\right)\right\}, \tag{5.19}$$

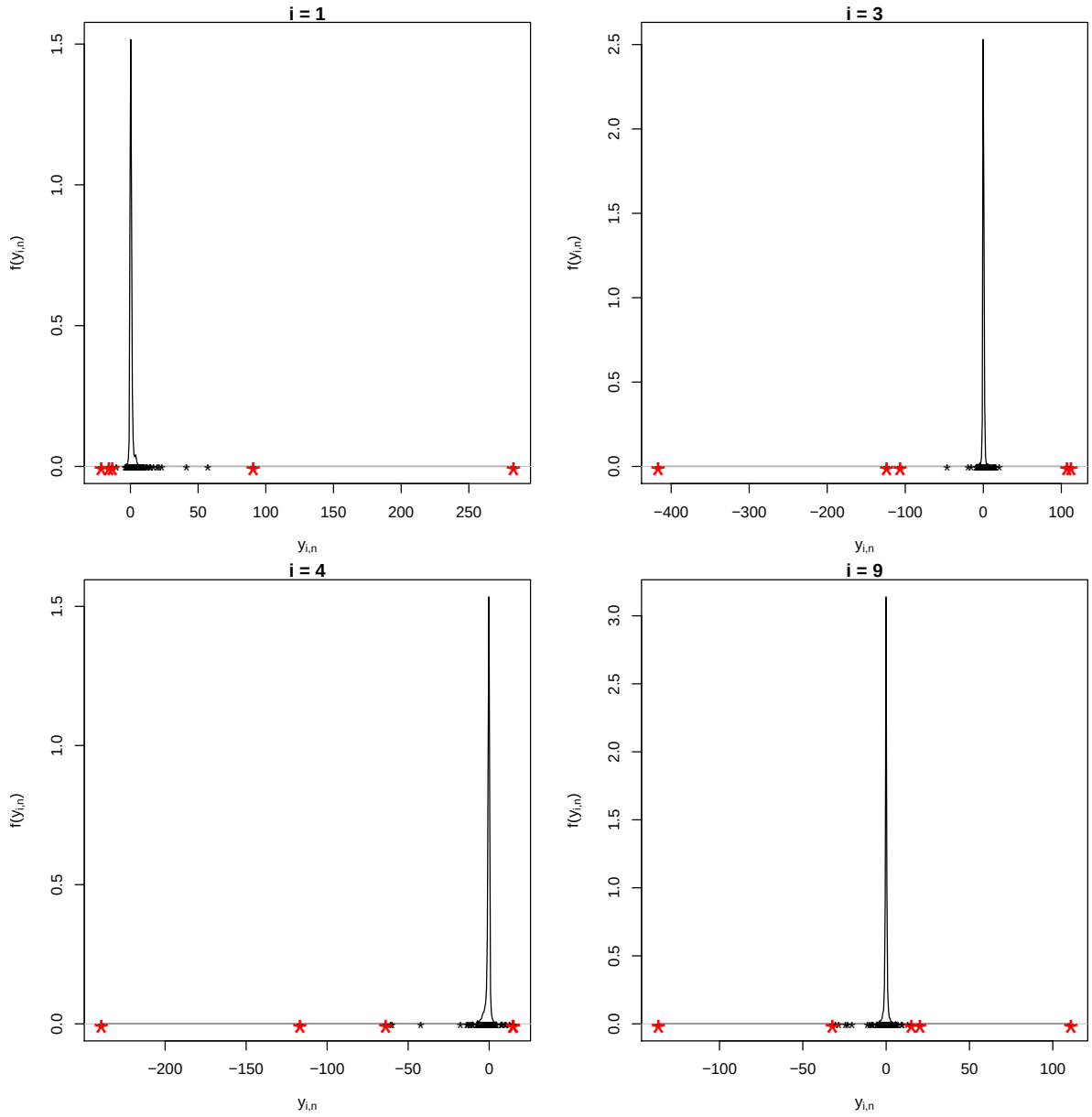


Figure 5.5: Plot of the kernel density estimates of realizations of $\boldsymbol{Y}_{v_i,n}$ for four specific random projections v_i, where the sample $\{x_1, \ldots, x_n\}$ used in $\boldsymbol{Y}_{v_i,n}$ (c.f. (5.16)) are realizations of a Brown-Resnick process (i.e. a simple max-stable process) and v_i are realizations of a Gaussian process. The points marked in red describe the five most extreme points of the distribution. Depending on the sign of the Gaussian process, the distributions of $\boldsymbol{Y}_{v_i,n}$ seem to sometimes be heavy-tailed on only one side (e.g. $i = 4$) and sometimes on both sides (e.g. $i = 9$).

with the notation as above in Definition 5.13 and where the index α_i is estimated by

$$\hat{\alpha}_i = 1/\hat{\gamma}\Big(\big|\langle v_i, x_1\rangle\big|, \ldots, \big|\langle v_i, x_n\rangle\big|\Big), \tag{5.20}$$

for $i = 1, \ldots, k$.

Remark 5.17. The multivariate functional random projection depth with minimum can be modified analogously using the extreme value theory corrected random Tukey depth defined in (5.11). However, the main issue is that a definition of a multivariate max-stable process needs yet to be developed. As a result, a lack of data availability limits any application of such a modified depth function.

5.5 Simulation Study for Functional Data

Analogously to the simulation study carried out in Section 5.3, we now seek to compare the extreme value theory corrected functional random projection depth with minimum defined in (5.19) to both non-modified versions of the functional random projection depth by applying the three depth functions once to Brown-Resnick data and once to extremal-t data.

As in the multivariate case, in practice, a good value κ for the Hill estimator needs to be determined first. For this, we study the Hill plot of the Hill estimator of γ implied by 5.20 and apply it to the sample of absolute values of the sample $\boldsymbol{Y}_{v_i,n}$ defined in 5.16, for $n = 1\,000$ stochastic processes and for several projections $v_i \in \{v_1, \ldots, v_k\}$. The results are given in the Appendix in Figure 8.5 and in Figure 8.6.

We now apply the three mentioned depth functions to both data sets with $k = 1\,000$ random projections to plot the obtained depth values pairwise against each other in DD-plots. The results are shown in Figure 5.6 for the for Brown Resnick data and in Figure 5.7 for the extremal-t data.

Both figures show similar results. We notice strong signs of degeneracy for the functional random projection depth with minimum, as already observed in previous simulations. However these seem to be alleviated by the corresponding extreme value theory corrected version. We also notice that this corrected depth function seems to correlate strongly with the functional random projection depth with mean in the tail. Finally, we revise the outlier detection study carried out at the end of Section 4.3 and add the extreme value theory corrected functional random projection depth with minimum to it to obtain, respectively, Figure 5.6 for the Brown-Resnick data study and Figure 5.7 for the extremal-t study. Our findings from the previous figures seem to be confirmed here, as the extreme value theory corrected depth function appears to successfully detect outliers, which is not the case for the non-modified version. However, given the good performance of the functional random projection depth with mean and its noticed correlation in the tail with the extreme value theory corrected functional random projection depth with minimum, it is questionable if this depth modification is worth the effort, in particular in light of the many made assumptions.

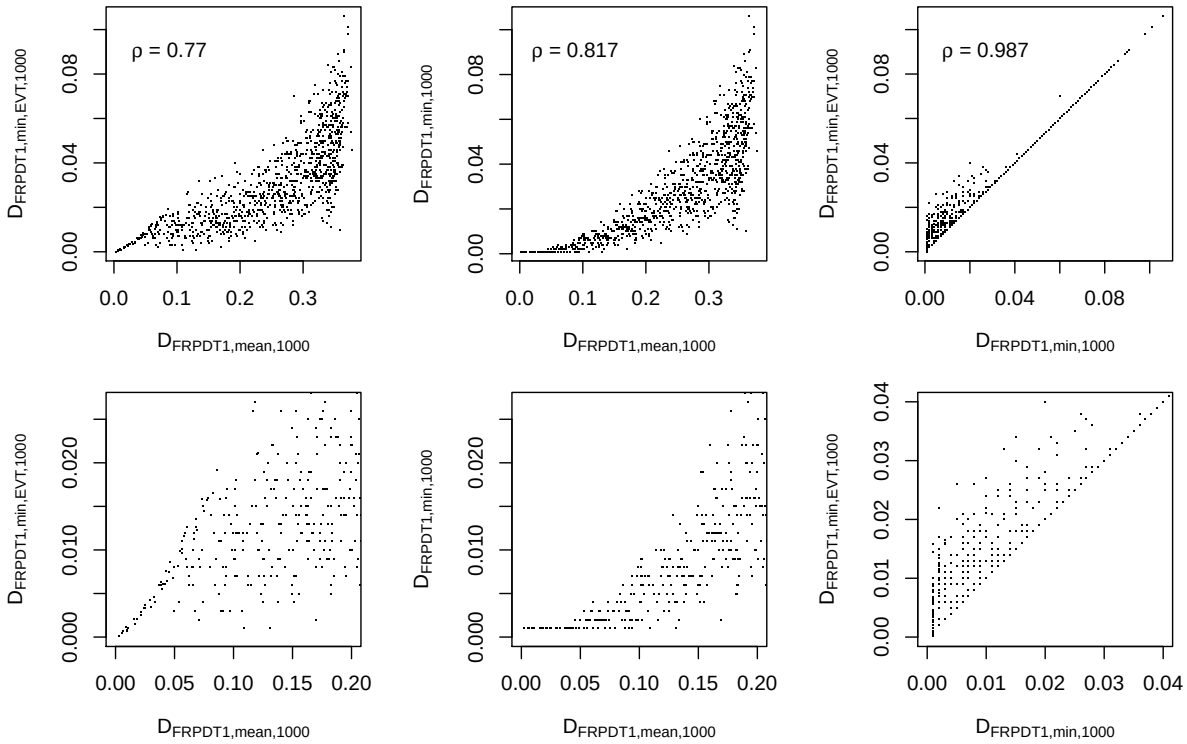


Figure 5.6: Pairwise DD-plots between the extreme value theory corrected functional random projection depth with minimum ($D_{FRPD,EVT,\min}$), the functional random projection depth with mean ($D_{FRPD,\text{mean}}$) and the functional random projection depth with minimum ($D_{FRPD,\min}$) for $n = 1\,000$ realizations of Brown-Resnick processes and $k = 1\,000$ random projections. The top three plots show full DD-plots, while the lower plots show a corresponding version that is zoomed to the lower left corner, i.e. where depth values are usually small for both depth functions.

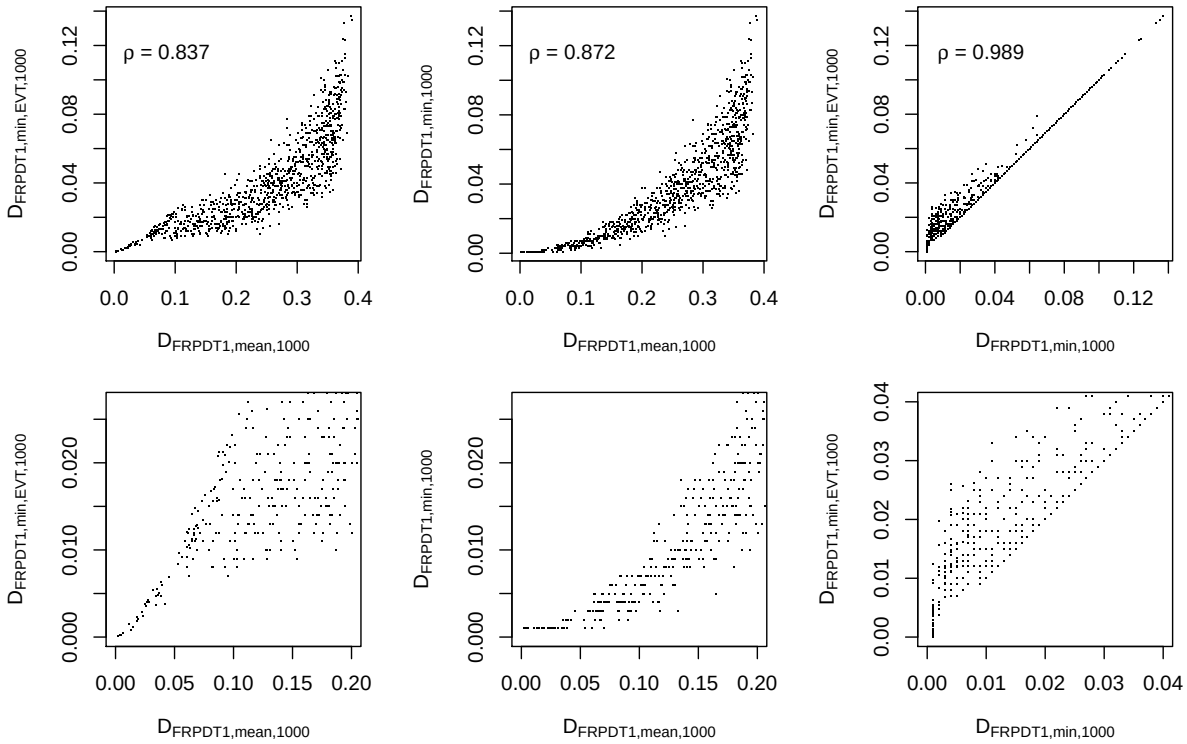


Figure 5.7: Pairwise DD-plots between the extreme value theory corrected functional random projection depth with minimum ($D_{FRPD,EVT,\min}$), the functional random projection depth with mean ($D_{FRPD,\text{mean}}$) and the functional random projection depth with minimum ($D_{FRPD,\min}$) for $n = 1\,000$ realizations of extremal-t processes and $k = 1\,000$ random projections. The top three plots show full DD-plots, while the lower plots show a corresponding version that is zoomed to the lower left corner, i.e. where depth values are usually small for both depth functions.

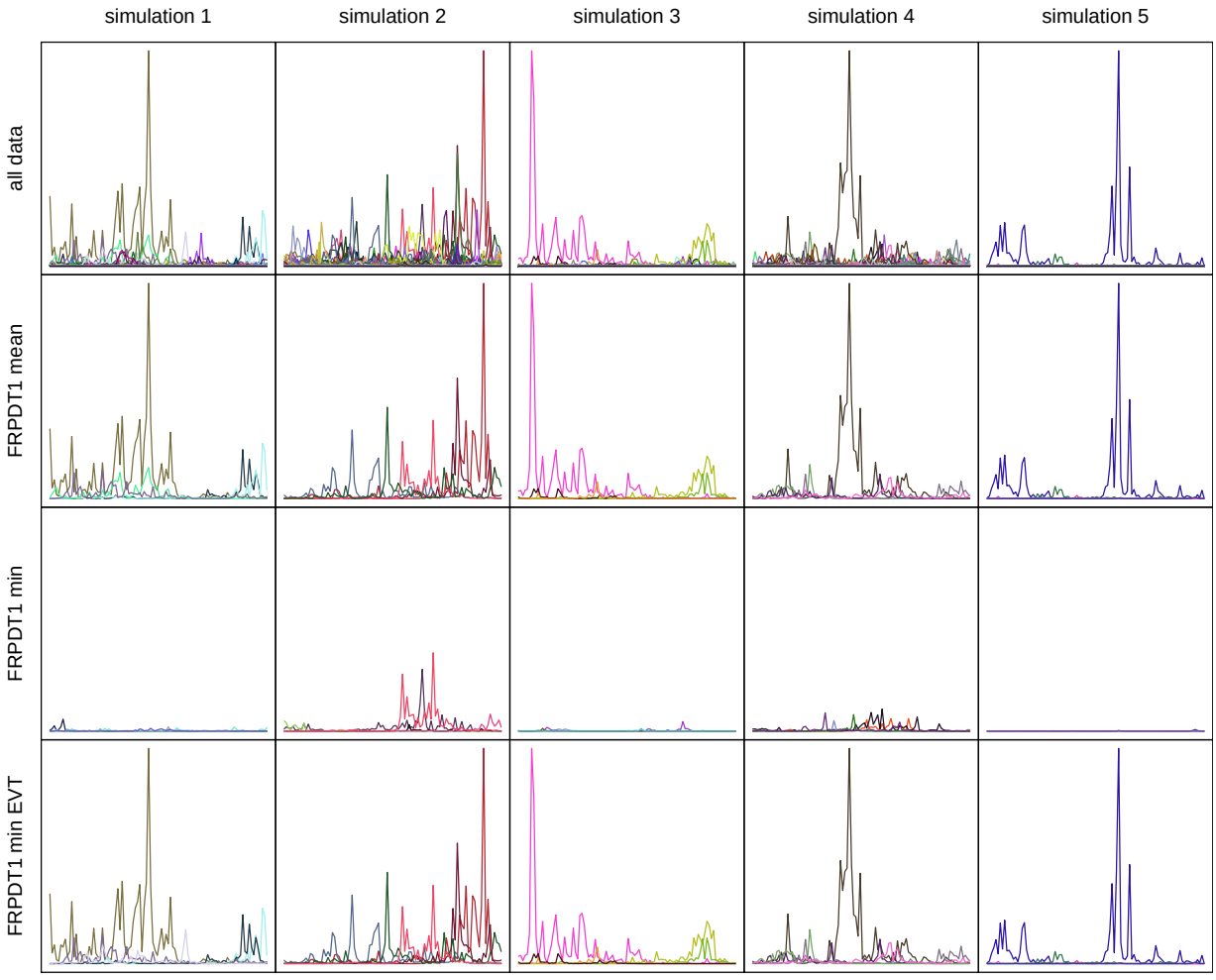


Figure 5.8: Plots of five $n = 1\,000$ Brown-Resnick data simulations (top row) and, respectively, plots of the five most extreme observations based on $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and $D_{FRPDT1,EVT,\text{min}}$ calculations (each following row relates to one depth function). Colors are chosen randomly, though consistency is kept across plots related to the same data.

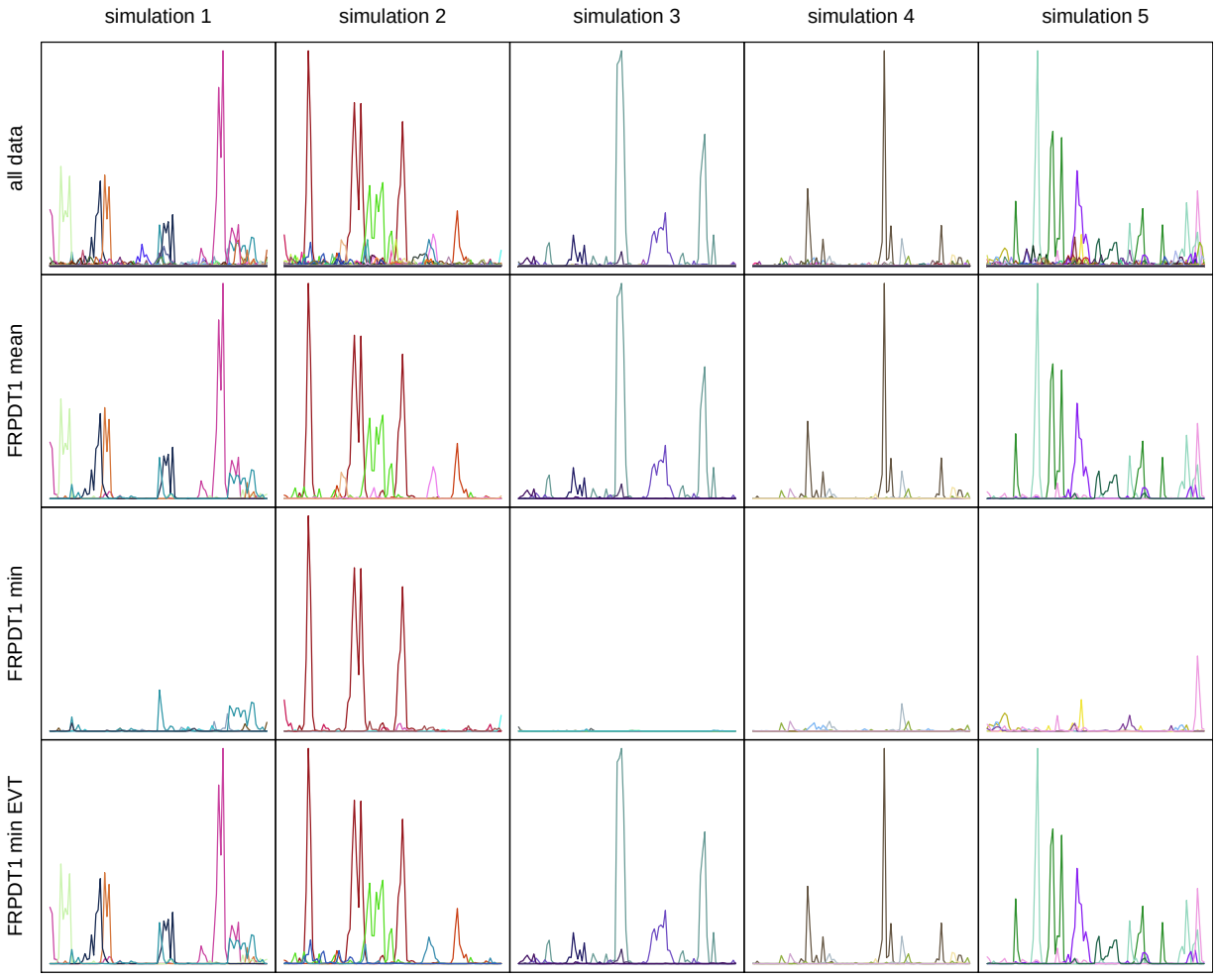


Figure 5.9: Plots of five $n = 1\,000$ extremal-t data simulations (top row) and, respectively, plots of the five most extreme observations based on $D_{FRPDT1,\text{mean}}$, $D_{FRPDT1,\text{min}}$ and $D_{FRPDT1,EVT,\text{min}}$ calculations (each following row relates to one depth function). Colors are chosen randomly, though consistency is kept across plots related to the same data.

6 Summary and Outlook

Having considered over a dozen existing functional depth functions, after removing special cases and non-performing depth functions, the number of depth functions showing strong potential is successfully reduced. For the univariate functional case, these depth functions are the integrated depth with additional weight function (Fraiman and Muniz, 2001), the functional random projection depth with mean (Cuevas et al., 2007) and the extremal depth (Narisetty and Nair, 2015). For the integrated depth, an additional interesting decomposition is presented by the total variation depth (Huang and Sun, 2016) when $D_M = D_{S1}$ is chosen. However, for the multivariate functional case, only the multivariate integrated depth (Claeskens et al., 2014) and the multivariate functional random projection depth with mean (Cuevas et al., 2007) are interesting. Five desirable properties for functional depth functions are identified and may be described as affine invariance, maximality at center, monotonicity relative to deepest point, null at the boundary and receptivity to convex hull width across the domain (Property P1 to Property P5). For the above mentioned functional depth functions, the property-wise performance is summarized in Table 6.1 below.

	P1	P2	P3	P4	P5
Univariate functional case					
Integrated depth	✓	✓	✓	✗	✓
Extremal depth	✓	✓	✓	✓	✗
FRPD with mean	✓	✓	✓	✓	✓
Multivariate functional case					
Multivariate integrated depth	✓	✓	✓	✗	✓
Multivariate FRPD with mean	✓	✓	✓	✓	✓

Table 6.1: Summary of properties for introduced functional depth functions showing strong potential. The symbol ✓ indicates that the property is satisfied and the symbol ✗ indicates that it is not. FRPD stands for "functional random projection depth".

The main drawback of the integrated depth and the multivariate integrated depth is that Property P4 - the null at the boundary property - is not satisfied. This mainly causes these depth functions to perform poorly for outlier detections as shown in Narisetty and Nair (2015), in Huang and Sun (2016) and in this thesis (see for example Figure 4.11 and Figure 4.12).

The extremal depth seems to be interesting and appears to have a strong potential for development as it satisfies all first four properties. One improvement could be to alter the depth function to reflect actual depth values instead of the ranks it currently calculates. For this, maybe results from

survival analysis could be used as an inspiration as the plots of the functions $\Phi_{x_i,n}, i = 1, \ldots, n$ in the left side of Figure 3.15 resemble (mirrored) survival functions and their utilization in the extremal depth construction shares commonalities with survival analysis. A second improvement could be a generalization of the extremal depth to the multivariate functional case and a third improvement could be to adapt the depth function to potentially satisfy Property P5.

The total variation depth decomposition of the integrated depth - although not considered in detail in this thesis - seems to show promising results that are worth further study, in particular in light of a potentially multivariate generalization. The decomposition allows for a successful detection of amplitude and shape outliers, with, as suggested in Huang and Sun (2016), better accuracy than the integrated depth without decomposition and than the extremal depth.

The functional random projection depth with mean and its multivariate generalization seem to be the most interesting depth functions as they are the only depth functions for functional data, that satisfy all five properties. Surprisingly, these depth functions have seen only very little attention in the literature so far.

The receptivity to convex hull width across the domain property - Property P5 - is only informally defined in this thesis and would require a formal definition, perhaps inspired by the weight function w_3 provided in the multivariate integrated depth in (3.37). However, this is a difficult task, as for example in Nagy et al. (2016), this weight function is explicitly avoided as deemed theoretically too difficult to deal with. The version of Property P5 presented in Nieto-Reyes and Battey (2014) is limited in its application, as the interval U in that definition is only divided in two intervals, one with high variability and one with low variability.

In general, all the above mentioned functional depth functions rely on depth functions for multivariate data in their construction and thus, these also need to be considered carefully. In particular, it is required that the depth functions for multivariate data satisfy Property M1 to Property M4 in order for the functional depth functions to satisfy Property P1 to Property P4. For a list of currently available depth functions for multivariate data, we refer to the **R** package "ddalpha" and the corresponding paper (Pokotylo, Mozharovskyi, and Dyckerhoff, 2016).

In order to apply extreme value theory results to functional depth functions, we first study two depth functions for multivariate data, the random Tukey depth (Cuesta-Albertos and Nieto-Reyes, 2007) and the random projection depth (Cuevas et al., 2007)[1]. The first result obtained is that the number of random projections required for the random Tukey depth is significantly higher than the number recommended in Cuesta-Albertos and Nieto-Reyes (2007). Instead of the suggested ten projections for dimensions up to $K = 50$, we find that for dimensions $K = 2$ to $K = 4$, between 20 and 800 projections are required. For higher dimensions, this number increases rapidly.

We then consider the extreme value theory modification of the random Tukey depth proposed in Einmahl et al. (2015) and find that this depth function seems to correlate strongly with the random projection depth. This is surprising as although the two depth functions are constructed similarly, the former depth function uses a minimum while the latter depth function uses a mean function in their

[1]Note that this random projection depth is different from the one presented in Pokotylo et al. (2016).

respective calculations. Here, studying the depth contour levels of the random projection depth, as for example done in He and Einmahl (2017), could lead to further insights into this very interesting similarity.

Unfortunately, all mentioned functional depth functions in Table 6.1, due to their design, are virtually not affected by an extreme value theory based correction. However, a modification for the functional random projection depth with minimum (Cuevas et al., 2007) seems to at least alleviate some of the degeneracies observed in the carried out simulations. For this, we use for the marginal depth the already studied extreme value theory corrected Tukey depth. When applying the functional random projection depth to simple max-stable processes, we assume that the inner products of the trajectories of simple max-stable and Gaussian processes are heavy-tailed. These assumptions seem to be confirmed by our simulations, although theoretical results would be preferable. From further simulations, it seems that the functional random projection depth with minimum and extreme value theory modification correlates with the functional random projection depth, in particular in the tail. While some further research might answer a few unanswered questions, we are not optimistic on finding any helpful extreme value theory modification for the current available depth functions for functional data.

7 Notation

Multivariate Case

X	random variable
x	realization or number
$\{x_1, , \ldots, x_n\}$	random sample, $n \in \mathbb{N}$
$\boldsymbol{X} = (X^{(1)}, \ldots, X^{(K)})$	random vector in $\mathbb{R}^K$, $K \in \mathbb{N}$
$\boldsymbol{x} = (x^{(1)}, \ldots, x^{(K)})$	realization or vector in $\mathbb{R}^K$, $K \in \mathbb{N}$
$\{\boldsymbol{x}_1, , \ldots, \boldsymbol{x}_n\}$	random K-variate sample, $n \in \mathbb{N}$

Functional Case

X	stochastic process
x	functional realization or function
$\{X_1, , \ldots, X_n\}$	sample of stochastic processes, $n \in \mathbb{N}$
$\boldsymbol{X} = (X^{(1)}, \ldots, X^{(K)})$	K-variate stochastic process, $K \in \mathbb{N}$
$\boldsymbol{x} = (x^{(1)}, \ldots, x^{(K)})$	K-variate functional realization or K-variate function, $K \in \mathbb{N}$
$\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$	sample of K-variate stochastic processes, $n \in \mathbb{N}, K \in \mathbb{N}$
$U \subseteq \mathbb{R}$	compact interval on which the stochastic process are evaluated
$\boldsymbol{X}(t) = \left(X^{(1)}(t), \ldots, X^{(K)}(t)\right) \in \mathbb{R}^K$	stochastic process evaluated at $t \in U$
$\boldsymbol{x}(t) = \left(x^{(1)}(t), \ldots, x^{(K)}(t)\right) \in \mathbb{R}^K$	function value evaluated at $t \in U$

General Notation

P	distribution of X or $\boldsymbol{X}$
F	cumulative distribution function of X or $\boldsymbol{X}$
$\boldsymbol{A}$	matrix
$\mathcal{P}$	class of distribution on the Borel sets
P_n	empirical distribution
$P_{t,n}$	marginal distribution of P_n at $t \in U$

F_n	empirical cumulative distribution function of the sample $\{x_1, \ldots, x_n\}$ or $\{\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}\}$
$\mathcal{C}(U)$	space of continuous functions on U
$\mathcal{L}^2(U)$	space of square integrable functions on U
$\mathcal{F}$	space of square integrable functions intersected with the space of continuous functions on U
$\mathcal{C}(U)^K$	space of K-variate continuous functions intersected with the space of K-variate continuous functions on U
$\mathcal{L}^2(U)^K$	space of K-variate square integrable functions on U
$\mathcal{F}_K$	space of K-variate square integrable functions intersected with the space of K-variate square integrable functions on U
Θ	center of symmetry of the distribution P (if it exists)
Σ	covariance matrix

Depth Functions

For Univariate and Multivariate Data

D_{T1}	Tukey depth for $K = 1$
D_T	Tukey depth for $K \geq 2$
$D_{RT,k,\nu}$	random Tukey depth with k realizations of random vectors chosen with probability measure ν
$D_{RP,k,\nu}$	random projection depth with k realizations of random vectors chosen with probability measure ν
$D_{RT,EVT,k,\nu}$	random Tukey depth with k realizations of random vectors chosen with probability measure ν and extreme value theory modification
D_{S1}	simplicial depth for $K = 1$
D_S	simplicial depth
D_{MH}	Mahalanobis depth
D_M	marginal depth for the corresponding general functional depth

For Functional Data

D_{ID}	integrated depth
$D_{FRPD,k,\eta,\text{mean}}$	functional random projection depth with k realizations of stochastic processes drawn with probability measure η, calculated with averages. $D_{FRPD,k,\eta} = D_{FRPD,k,\eta,\text{mean}}$, when not specified.
$D_{FRPD,k,\eta,\min}$	functional random projection depth with k realizations of stochastic processes drawn with probability measure η, calculated with minimum

$D_{FRPD,EVT,k,\eta,\min}$	functional random projection depth with k realizations of stochastic processes drawn with probability measure η, calculated with minimum and with D_M chosen as the extreme value theory corrected Tukey depth.
D_{HRD}	half-region depth
D_{MHRD}	modified half-region depth
D_{BD}	band depth
D_{MBD_J}	modified band depth with J-bands
D_{ED}	extremal depth
D_{TVD}	total variation depth
D_{MID}	multivariate integrated depth
D_{MID_α}	multivariate integrated depth with weight function w_3
D_{WMID}	weighted multivariate integrated depth
$D_{MFRPD,k,\eta,\text{mean}}$	multivariate functional random projection depth with k realizations of stochastic processes drawn with probability measure η, calculated with averages. $D_{MFRPD,k,\eta} = D_{MFRPD,k,\eta,\text{mean}}$, when not specified.
$D_{MFRPD,k,\eta,\min}$	multivariate functional random projection depth with k realizations of stochastic processes drawn with probability measure η, calculated with minimum
D_{SBD}	simplicial band depth
D_{MSBD}	modified simplicial band depth
D_{XYZW}	functional depth D_{XY} with marginal depth $D_M = D_{ZW}$

Extreme Value Theory

α	tail index
γ	extreme value index
$\hat{\gamma}$	Hill estimator for γ
DoA	domain of attraction
G_γ	extreme value distribution in $\mathbb{R}$
$F_{n,EVT}$	extreme value theory corrected cumulative distribution function
$F_{\boldsymbol{v},n,EVT}$	marginal extreme value theory corrected cumulative distribution function
p_r	right tail probability
$p_{r,n}$	estimator for the right tail probability
p_l	left tail probability
$p_{l,n}$	estimator for the left tail probability
RV_α	regularly varying with index α

8 Appendix

8.1 Gaussian Process Regression for Simulations

To simulate the kind of Gaussian processes used in thesis, we use concepts from *Gaussian process regression* introduced in Ebden (2015). Contrarily to a linear regression, where we expect a function to be of the form $f(t) = mt + c$, with m and c constants in $\mathbb{R}$, $t \in U$, a Gaussian process regression models the function $f(t)$ as a Gaussian process. More precisely, since $f(t)$ is usually only observed for finite points $t_1, \ldots, t_{T_1} \in U$ we can model the sample $\{y_1, \ldots, y_{T_1}\} := \{f(t_1), \ldots, f(t_{T_1})\}$ as a multivariate normal distribution. Often it is assumed that this distribution has mean zero and what relates one observation to another is only the covariance function $\Sigma(t_i, t_j)$, $t_i, t_j \in U$. A popular choice for this covariance function is the "squared exponential" given by

$$\Sigma_1(t_i, t_j) := \sigma_f^2 \exp\left[\frac{-(t_i - t_j)^2}{2l^2}\right], \tag{8.1}$$

where $\sigma_f^2 > 0$ and $l > 0$ are parameters to fine tune the covariance function and $t_i, t_j \in U$. The parameter σ_f^2 determines the variability of the process and the parameter l is used to modify how strongly the observed points $f(t_i)$ and $f(t_j)$ are correlated with one another depending on the distance between t_i and t_j, $t_i, t_j \in U$.

Now, observing the data with noise we have

$$y = f(t) + \mathcal{N}(0, \sigma_\epsilon^2),$$

with $\sigma_\epsilon^2 > 0$. By incorporating the error term into the covariance function, we modify (8.1) to obtain

$$\Sigma_2(t_i, t_j) := \sigma_f^2 \exp\left[\frac{-(t_i - t_j)^2}{2l^2}\right] + \sigma_\epsilon^2 \delta(t_i, t_j), \tag{8.2}$$

where $\delta(t_i, t_j)$ is the Kronecker delta function defined by

$$\delta(t_i, t_j) := \begin{cases} 1 & \text{for } t_i = t_j \\ 0 & \text{for } t_i \neq t_j \end{cases}.$$

Given the sample $\boldsymbol{y} := (y_1, \ldots, y_{T_1})$, we want to predict the observation y_* corresponding to the

chosen time point t_*. Assuming the Gaussian process model, we have that

$$\begin{pmatrix} \boldsymbol{y} \\ y_* \end{pmatrix} \sim \mathcal{N}\left[\mathbf{0}, \begin{pmatrix} \boldsymbol{M} & \boldsymbol{M}_*' \\ \boldsymbol{M}_* & M_{**} \end{pmatrix}\right],$$

where M_*' is the transposed matrix of M_* and

$$\boldsymbol{M} := \begin{pmatrix} \Sigma_2(t_1,t_1) & \Sigma_2(t_1,t_2) & \dots & \Sigma_2(t_1,t_{T_1}) \\ \Sigma_2(t_2,t_1) & \Sigma_2(t_2,t_2) & \dots & \Sigma_2(t_2,t_{T_1}) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_2(t_{T_1},t_1) & \Sigma_2(t_{T_1},t_2) & \dots & \Sigma_2(t_{T_1},t_{T_1}) \end{pmatrix}$$

$$\boldsymbol{M}_* := \Big(\Sigma_2(t_*,t_1), \Sigma_2(t_*,t_2), \dots, \Sigma_2(t_*,t_{T_1})\Big) \quad M_{**} := \Big(\Sigma_2(t_*,t_*)\Big). \tag{8.3}$$

Thus, the conditional probability of y_* given the data leads to the following distribution

$$y_*|\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{y}, M_{**} - \boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{M}_*'),$$

i.e. $y_*|\boldsymbol{y}$ has mean $\boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{y}$ and variance $(M_{**} - \boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{K}_*')$. The generalization of this distribution to T_2 number of predictions follows naturally by increasing the dimensions of $\boldsymbol{M}_*$ and M_{**} in (8.3). This leads to

$$\boldsymbol{y}_*|\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{y}, \boldsymbol{M}_{**} - \boldsymbol{M}_*\boldsymbol{M}^{-1}\boldsymbol{M}_*'), \tag{8.4}$$

where $\boldsymbol{y}_* \in \mathbb{R}^{T_2}, \boldsymbol{y} \in \mathbb{R}^{T_1}, \boldsymbol{M} \in \mathbb{R}^{T_1 \times T_1}, \boldsymbol{M}_* \in \mathbb{R}^{T_2 \times T_1}, \boldsymbol{M}_{**} \in \mathbb{R}^{T_2 \times T_2}$.

We now use this distribution to simulate "smoothed" Gaussian processes. Set for example $\sigma_f^2 := 1$, $l := 1$, $\sigma_\epsilon^2 := 0$,

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \end{pmatrix} := \begin{pmatrix} 0 \\ 4 \\ 7 \\ 10 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} f(t_1) \\ f(t_2) \\ f(t_3) \\ f(t_4) \end{pmatrix} := \begin{pmatrix} -1 \\ 0 \\ 3 \\ 0 \end{pmatrix}.$$

To simulate $n = 50$ Gaussian processes on the interval $U = [0, 10]$ discretized on $T := T_1 + T_2 = 4 + 96 = 100$ number of points we draw 50 times from (8.4) and obtain the data presented in Figure 8.1.

Further changing the value of the variance of the error from $\sigma_\epsilon^2 = 0$ to $\sigma_\epsilon^2 = 2$ yields Figure 8.2. Note that in the simulations σ_f^2 determines the variability between t_i and t_{i+1}, $t_i, t_{i+1} \in \{t_1, \dots, t_{T_1}\}$, while σ_ϵ^2 determines a global variability.

This approach is used to simulate the data seen in Figure 1.1, Figure 2.2, Figure 2.3, Figure 3.7, Figure 3.8, Figure 3.15 (top left), Figure 4.2, Figure 4.3 and Figure 4.7.

To simulate Gaussian processes with varying amplitude, we change the constant variance of the error σ_ϵ^2 used in (8.2) into a covariance function on $t_i \in U$. I.e.

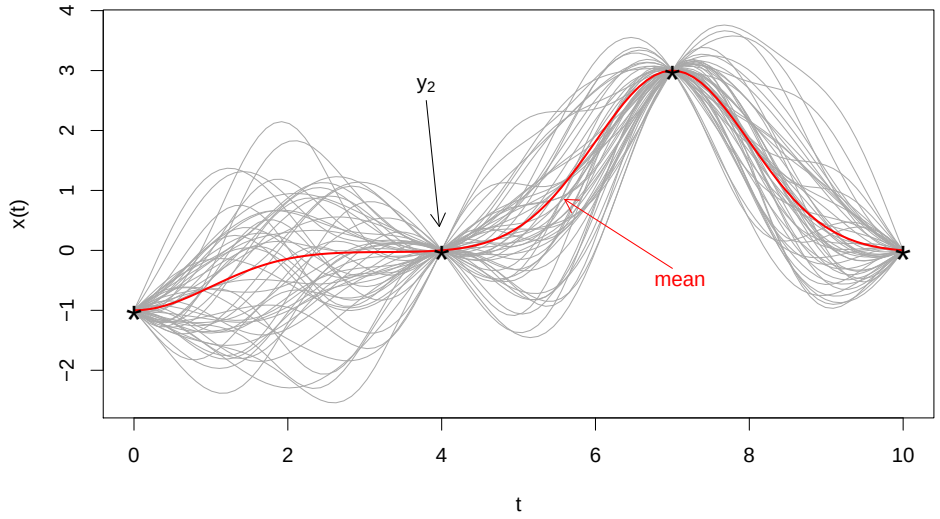


Figure 8.1: Illustration of $n = 50$ realizations of Gaussian processes simulated with $T = 100$ discretization points and four fixed values (black points). Further, $l = 1$, $\sigma_f^2 = 0.4$, $\sigma_\epsilon^2 = 0$ and the expected mean is drawn in red.

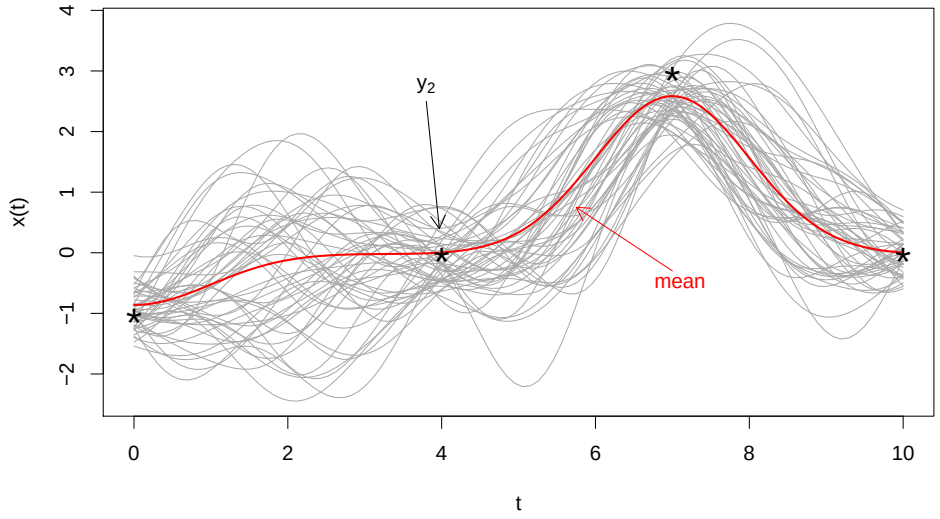


Figure 8.2: Illustration of $n = 50$ realizations of Gaussian simulated in with the same settings as in Figure 8.1, except with $\sigma_\epsilon^2 = 0.4$.

$$\Sigma_3(t_i, t_j) := \sigma_f^2 \exp\left[\frac{-(t_i - t_j)^2}{2l^2}\right] + \sigma_\epsilon^2(t_i)\delta(t_i, t_j),$$

where $\sigma_\epsilon^2(t_i) > 0$ is some function on $\mathcal{C}(U)$, $t_i \in U$. By experimenting with this function, we obtain the data used in Figure 2.4, Figure 4.2 (top right), Figure 4.4 and Figure 4.8.

8.2 Simple Max-stable Processes

The second type of data sets used in this thesis are data generated from simple max-stable processes. Two such processes considered are the Brown-Resnick process introduced in Brown and Resnick (1977) and the extremal-t process introduced in Opitz (2012).

Let $\mathscr{X}$ be a compact subset of $\mathbb{R}^K$, $K \in \mathbb{N}$. A simple max-stable process, defined in Definition 5.10, is known to have a spectral characterization for which a representation is given in Schlather (2002). Let $\{\zeta_i : i \in \mathbb{N}\}$ be the points of a Poisson process on $(0, \infty)$ with intensity $d\Lambda(\zeta) = \zeta^{-2}d\zeta$. Then, it is shown that there exists a non-negative stochastic process Y with continuous sample paths on $\mathscr{X}$ such that $\mathbb{E}\{Y(x)\} = 1$ for all $x \in \mathbb{R}^K$ and for which the limiting distribution G in (5.9) has the same distribution as

$$\max_{i \geq 1} \zeta_i Y_i(x), \qquad x \in \mathscr{X}, \tag{8.5}$$

where Y_i are independent replications of Y.

With this, we can define the Brown-Resnick model and the extremal-t model as follows.

Definition 8.1. The Brown-Resnick model (Brown and Resnick, 1977) is given by taking $Y_i(x) = \exp\{\xi_i(x) - \sigma^2(x)/2\}$ in (8.5), where ξ_i are independent replications of a centered Gaussian process with stationary increments such that $\mathrm{Var}\big[Y(x)\big] = \sigma^2(x)$ for all $x \in \mathscr{X}$. Its bivariate cumulative distribution function is given by

$$P\big(Z(x_1) \leq z_1, Z(x_2) \leq z_2\big) = \exp\left[-\frac{1}{z_1}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\left(\frac{z_2}{z_1}\right)\right) - \frac{1}{z_2}\Phi\left(\frac{a}{2} + \frac{1}{a}\log\left(\frac{z_1}{z_2}\right)\right)\right],$$

where Φ denotes the standard normal cumulative distribution function and with $a^2 = \mathrm{Var}\big[Y(x_1 - x_2)\big]$.

Definition 8.2. The extremal-t model (Opitz, 2012) is given by taking

$$Y_i(x) = c_\nu \max\{0, \xi_i(x)\}^\nu, \qquad c_\nu = \sqrt{\pi}2^{-(\nu-2)/2}\Gamma\left(\frac{\nu+1}{2}\right)^{-1}, \quad \nu \leq 1,$$

where ξ is a standard Gaussian process with correlation function ρ and Γ is the Gamma function. Its bivariate cumulative distribution function is given by

$$P\big(Z(x_1) \leq z_1, Z(x_2) \leq z_2\big) = \exp\left[-\frac{1}{z_1} T_{\nu+1}\left\{ -\frac{\rho(x_1 - x_2)}{b} + \frac{1}{b}\left(\frac{z_2}{z_1}\right)^{1/\nu} \right\} - \frac{1}{z_2} T_{\nu+1}\left\{ -\frac{\rho(x_1 - x_2)}{b} + \frac{1}{b}\left(\frac{z_1}{z_2}\right)^{1/\nu} \right\} \right],$$

where T_ν is the cumulative distribution function of a Student random variable with ν degrees of freedom and $b^2 = \big(1 - \rho(x_1 - x_2)^2\big)/(\nu + 1)$.

For further reading about the theory of these processes and their simulations, we refer to Ribatet (2013), Davison et al. (2012) and the corresponding **R** package "SpatialExtremes".

8.3 Additional Plots

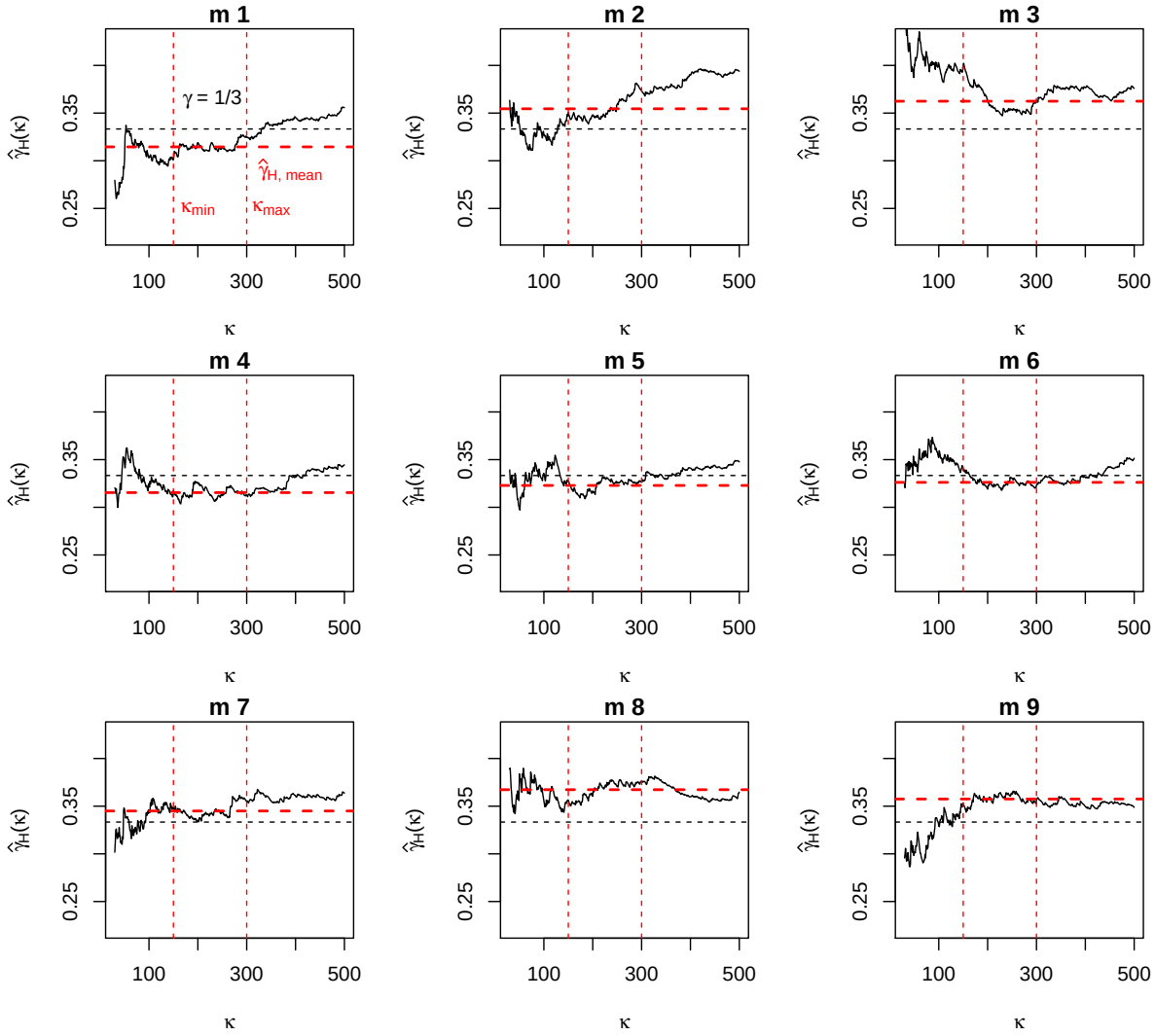


Figure 8.3: For Section 5.3: Hill plots for nine Monte Carlo simulations, each with $n = 1\,000$ realizations of data from the introduced clover distribution in ... . The interval, where variability is low over all plots is estimated to be $[150, 300]$. On each plot, the horizontal dashed red line indicates the average Hill estimator over this interval and the horizontal dashed black line indicates the true value of gamma, here $\gamma = 1/3$.

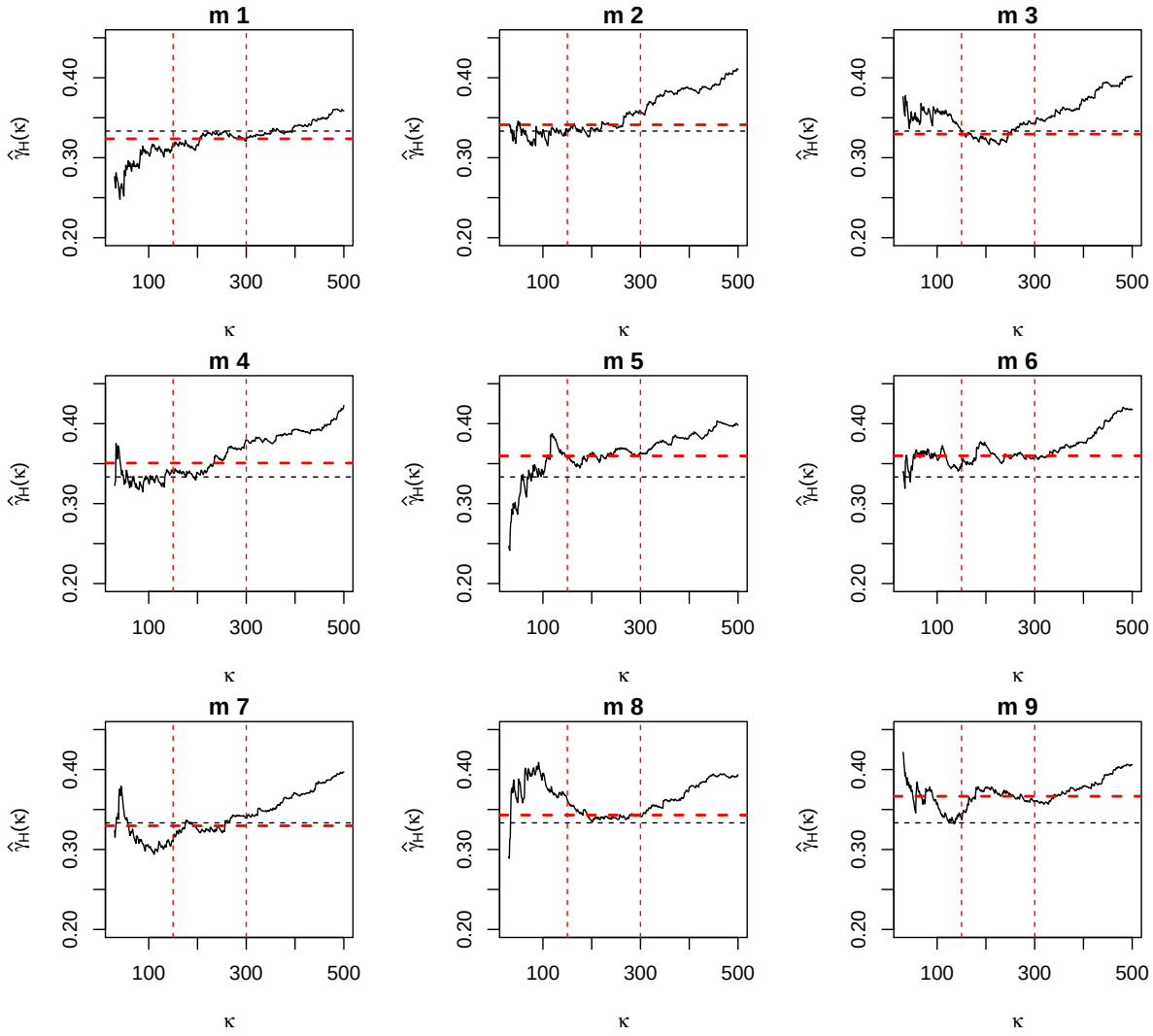


Figure 8.4: For Section 5.3: Hill plots for nine Monte Carlo simulations, each with $n = 1\,000$ realizations of data from the introduced elliptical distribution in ... . The interval, where variability is low over all plots is estimated to be $[150, 300]$. On each plot, the horizontal dashed red line indicates the average Hill estimator over this interval and the horizontal dashed black line indicates the true value of gamma, here $\gamma = 1/3$.

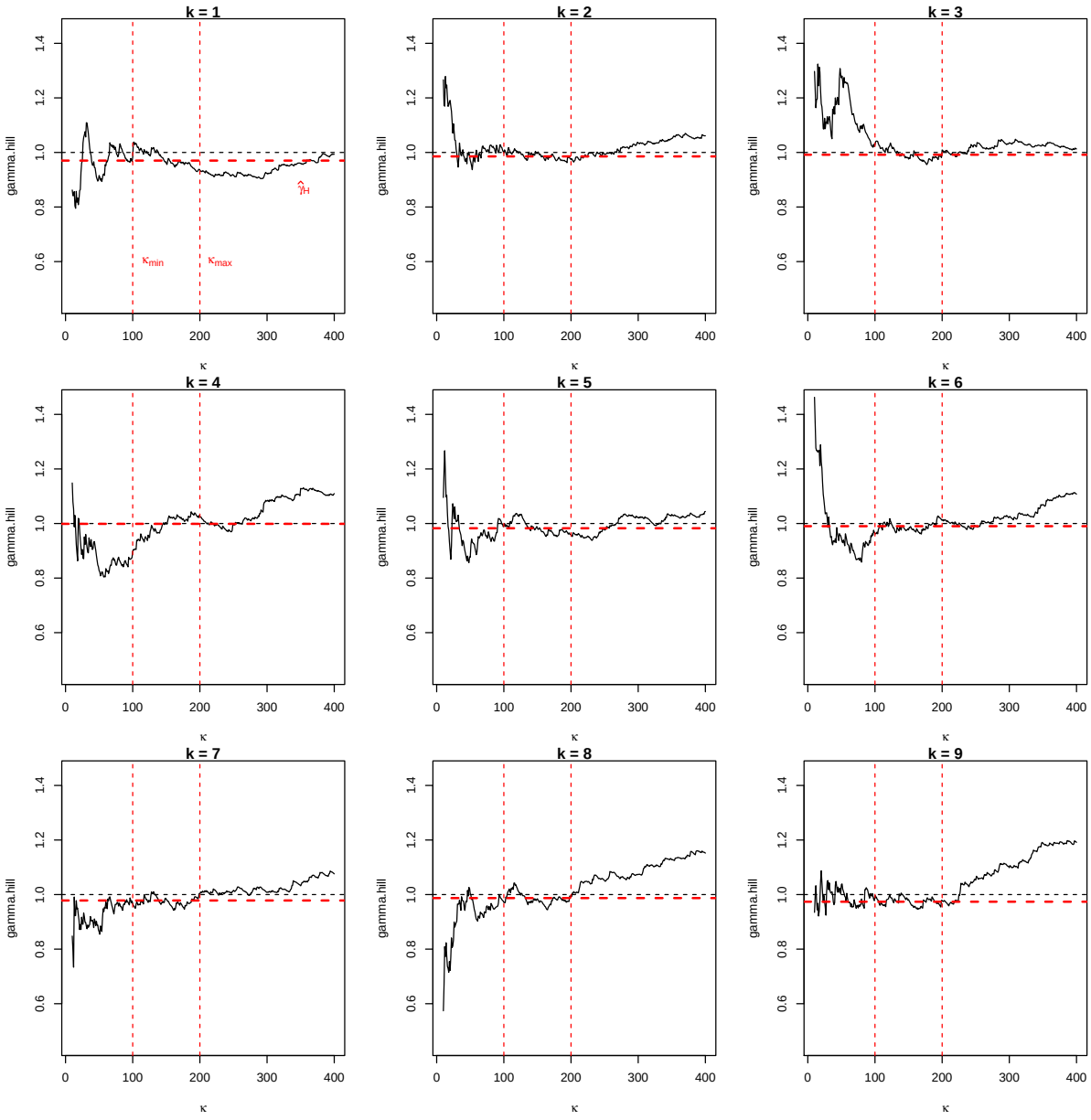


Figure 8.5: For Section 5.5: Nine Hill plots for projections $v_1, \ldots, v_9$ onto a realization of a Brown-Resnick process. The intervals $[100, 200]$ show low variability over all nine plots. The horizontal dotted red line indicates the average Hill estimator over this interval on each plot. For further calculation, a value of $\kappa = 150$ is used.

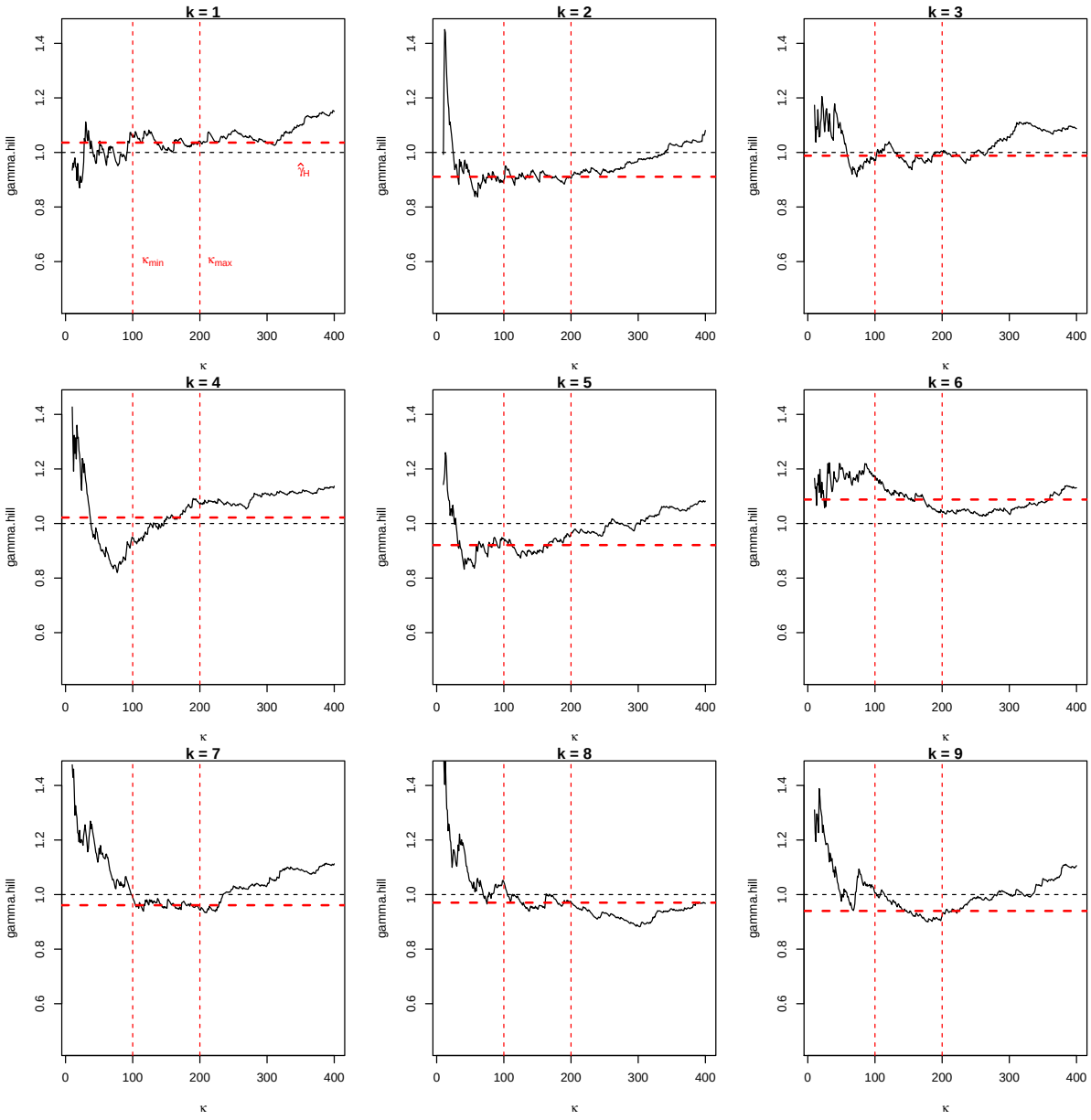


Figure 8.6: For Section 5.5: Nine Hill plots for projections $v_1, \ldots, v_9$ onto a realization of a extremal-t process. The intervals $[100, 200]$ show low variability over all nine plots. The horizontal dotted red line indicates the average Hill estimator over this interval on each plot. For further calculation, a value of $\kappa = 150$ is used.

8.4 R Code Overview

All **R** codes are made available on the following **GitHub** repository:
https://github.com/Gordon90s/R_codes_master_thesis

The content is:

1. Code **_RUN_FIRST_package_loader_directory_setting.R**
   - Loads and installs all the required **R** packages and implemented functions used throughout this thesis.
2. Code **univariate_functional_depths.R**
   - Implementation of the integrated depth with optional weight function w_3 and with optional extreme value theory modification.
   - Implementation of the extremal depth.
   - Implementation of the functional random projection depth with mean and minimum and optional extreme value theory modification for both.
   - Enables the use of the band depth, the modified band depth, the half-region depth and the modified half-region depth after installing the **R** package "ldfun" by Agostinelli (Agostinelli, 2015).
3. Code **EVT_functions.R**
   - Implementation of the moment and Hill estimators for the extreme value index γ, including the generation of the necessary Hill-plots.
   - Implementation of the right tail probability estimator.
   - Implementation of several extreme value theory corrected cumulative empirical distribution functions (among other, one and two sided versions).
   - Implementation of the univariate Tukey depth and the univariate simplicial depth.
   - Implementation of the multivariate random Tukey depth and the multivariate random projection depth both with and without extreme value theory modification.
   - Implementation of simulation functions for various heavy tailed bivariate distribution functions (obtained from He and Einmahl (2017)).
4. Code **data_simulation.R**
   - Simulation and saving of the four main data sets used in this thesis (i.e. the data sets presented in Figure 4.2).
5. Code **MT_MAIN_depth_function_for_analysis.R**

- Enables to use several depth functions developed in **univariate_functional_depths.R** all at once and in a standardized way. This function is primary used to generate the figures present in this thesis.

6. Folder **Comparaison_data_simulations**

   - For comparisons purposes and because simulation of max-stable processes take a long time, several data sets have been simulated and stored.

7. Code **secondary_functions.R**

   - Various secondary functions to mostly facilitate plotting.

8. Folder **R_Data_choice_of_k_RTD**

   - Contains all **R** codes used to determine the optimal value k_0 of random projections for the random Tukey depth, i.e. the simulations to obtain the values of Table 3.1. (Note that simulations take several days on standard computers.)

9. Folder **R_Data_choice_of_k_FRPD**

   - Contains all **R** codes used to determine the optimal value k_0 of functional random projections for the functional random projection depth, i.e. the simulations to obtain the values of Table 3.2. (Here too simulations take several days on standard computers.)

10. Folder **MT_plots**

    - Contains most of the **R** codes used to generate the figures present in this thesis. (For the remaining plots, the corresponding **R** codes are integrated in the simulations codes in the other above mentioned folders.)

Bibliography

Afshani, Peyman, Donald R. Sheehy, and Yannik Stein (2015), "Approximating the simplicial depth." *CoRR*, abs/1512.04856.

Agostinelli, Claudio (2015), "Local half-region depth for functional data."

Basrak, Bojan, Richard A. Davis, and Thomas Mikosch (2002), "A characterization of multivariate regular variation." *The Annals of Applied Probability*, 12, 908–920.

Billingsley, Patrick (1968), *Convergence of probability measures*. Wiley Series in probability and Mathematical Statistics: Tracts on probability and statistics, Wiley.

Billingsley, Patrick (2012), *Probability and Measure*. Wiley Series in Probability and Statistics, Wiley.

Brown, Bruce M. and Sidney I. Resnick (1977), "Extreme values of independent stochastic processes." *Journal of Applied Probability*, 14, 732–739.

Chakraborty, Anirvan and Probal Chaudhuri (2014), "The deepest point for distributions in infinite dimensional spaces."

Claeskens, Gerda, Mia Hubert, Leen Slaets, and Kaveh Vakili (2014), "Multivariate functional halfspace depth." *Journal of the American Statistical Association*, 109, 411–423.

Cuesta-Albertos, Juan. A. and Alicia Nieto-Reyes (2007), "The random Tukey depth." *ArXiv e-prints*.

Cuevas, Antonio, Manuel Febrero, and Ricardo Fraiman (2007), "Robust estimation and classification for functional data via projection-based depth notions." *Computational Statistics*, 22, 481–496.

Davison, Anthony. C., Simone A. Padoan, and Mathieu Ribatet (2012), "Statistical modeling of spatial extremes."

de Haan, Laurens and Ana Ferreira (2006), *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering, Springer New York.

Donoho, David L. and Miriam Gasko (1992), "Breakdown properties of location estimates based on halfspace depth and projected outlyingness." *Ann. Statist.*, 20, 1803–1827.

Dutta, Subhajit, Anil K. Ghosh, and Probal Chaudhuri (2012), "Some intriguing properties of Tukey's half-space depth."

Ebden, Mark (2015), "Gaussian processes: A quick introduction."

Einmahl, John H. J., Jun Li, and Regina Y. Liu (2015), "Bridging centrality and extremity: Refining empirical data depth using extreme value statistics." *ArXiv e-prints*.

Fraiman, Ricardo, Jean Meloche, Luis A. García-Escudero, Alfonso Gordaliza, Xuming He, Ricardo Maronna, Víctor J. Yohai, Simon J. Sheather, Joseph W. McKean, Christopher G. Small, Andrew Wood, R. Fraiman, and Jean Meloche (1999), "Multivariate L-estimation." *Test*, 8, 255–317.

Fraiman, Ricardo and Graciela Muniz (2001), "Trimmed means for functional data." *Test*, 10, 419–440.

Fubini, Guido (1907), "Sugli integrali multipli." *Rom. Acc. L. Rend. (5)*, 16, 608–614.

Ghosh, Anil K. and Probal Chaudhuri (2005), "On data depth and distribution-free discriminant analysis using separating surfaces." *Bernoulli*, 11, 1–27.

Gnedenko, Boris V. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire." *Annals of Mathematics*, 44, 423–453.

He, Yi and John H. J. Einmahl (2017), "Estimation of extreme depth-based quantile regions." *Journals of the Royal Statistical Society*, Series B. Statistical Methodology.

Hill, Bruce M. (1975), "A simple general approach to inference about the tail of a distribution." *Ann. Statist.*, 3, 1163–1174.

Hlubinka, Daniel, Irène Gijbels, Marek Omelka, and Stanislav Nagy (2015), "Integrated data depth for smooth functions and its application in supervised classification." *Computational Statistics*, 30, 1011–1031.

Huang, Huang and Ying Sun (2016), "Total Variation Depth for Functional Data." *ArXiv e-prints*.

Ieva, Francesca and Anna M. Paganoni (2013), "Depth measures for multivariate functional data." *Communications in Statistics: Theory and Methods*, 42, 1265–1276.

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

Li, Jun, Juan A. Cuesta-Albertos, and Regina Y. Liu (2012), "DD-classifier: Nonparametric classification procedure based on DD-plot." *Journal of the American Statistical Association*, 107, 737–753.

Liu, Regina Y. (1990), "On a notion of data depth based on random simplices." *The Annals of Statistics*, 18, 405–414.

Liu, Regina Y., Jesse M. Parelius, and Kesar Singh (1999), "Multivariate analysis by data depth: descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh)." *Ann. Statist.*, 27, 783–858.

Liu, Regina Y. and Kesar Singh (1997), "Notions of limiting p values based on data depth and bootstrap." *Journal of the American Statistical Association*, 92, 266–277.

López-Pintado, Sara and Juan Romo (2006), "Depth-based classification for functional data." *DIMACS Series in Discrete Mathematics and Theoretical Computer Science. Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications. American Mathematical Society.*, 72, 103–120.

López-Pintado, Sara and Juan Romo (2009), "On the concept of depth for functional data." *Journal of the American Statistical Association*, 104, 718–734.

López-Pintado, Sara and Juan Romo (2011), "A half-region depth for functional data." *Comput. Stat. Data Anal.*, 55, 1679–1695.

López-Pintado, Sara, Ying Sun, Juan K. Lin, and Marc G. Genton (2014), "Simplicial band depth for multivariate functional data." *Advances in Data Analysis and Classification*, 8, 321–338.

Mahalanobis, Prasanta C. (1936), "On the generalised distance in statistics." In *Proceedings National Institute of Science, India*, volume 2, 49–55.

Mosler, Karl and Yulia Polyakova (2012), "General notions of depth for functional data."

Mosler, Karl C. (2002), *Multivariate Dispersion, Central Regions, and Depth the Lift Zonoid Approach.* Lecture notes in statistics (Springer-Verlag), Springer New York :, New York, NY.

Nagy, Stanislav, Irène Gijbels, Marek Omelka, and Daniel Hlubinka (2016), "Integrated depth for functional data: statistical properties and consistency." *ESAIM: PS*, 20, 95–130.

Narisetty, Naveen N. and Vijayan N. Nair (2015), "Extremal Depth for Functional Data and Applications." *ArXiv e-prints.*

Nieto-Reyes, Alicia and Heather Battey (2014), "A topologically valid definition of depth for functional data." *ArXiv e-prints.*

Oja, Hannu (1983), "Descriptive statistics for multivariate distributions." *Statistics & Probability Letters*, 1, 327–332.

Opitz, Thomas (2012), "Extremal t processes: Elliptical domain of attraction and a spectral representation."

Pokotylo, Oleksii, Pavlo Mozharovskyi, and Rainer Dyckerhoff (2016), "Depth and depth-based classification with **R**-package ddalpha."

Resnick, Sidney I. (1987), *Extreme Values, Regular Variation and Point Processes*, first edition. Springer Series in Operations Research and Financial Engineering, Springer-Verlag New York.

Ribatet, Mathieu (2013), "Spatial extremes: Max-stable processes at work." *Journal de la Société Française de Statistique*, 2, 156–177.

Schlather, Martin (2002), "Models for stationary max-stable random fields." *Extremes*, 5, 33–44.

Tukey, John W. (1974), "Mathematics and the Picturing of Data." In *International Congress of Mathematicians 1974* (Ralph D. James, ed.), volume 2, 523–532.

Vardi, Yehuda and Cun-Hui Zhang (2000), "The multivariate L_1-median and associated data depth." *Proceedings of the National Academy of Sciences*, 97, 1423–1426.

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

Zuo, Yijun and Robert Serfling (2000), "General notions of statistical depth function." *Ann. Statist.*, 28, 461–482.