Session 7: Multivariate and functional depth
Winter course, CMStatistics 2016

Mia Hubert, Peter Rousseeuw, Stefan Van Aelst

Department of Mathematics
KU Leuven, Belgium

December 6–7, 2016

Outline of the course

- 1. General notions of robustness
- 2. Robustness for univariate data
- 3. Robust multivariate methods
- 4. Robust regression
- 5. Robust principal component analysis
- 6. Inference
- 7. Multivariate and functional depth
- 8. High dimensional data and sparsity
- 9. Cellwise outliers

Outline

1. Univariate data
2. Multivariate data
   - Halfspace depth and bagdistance
   - Projection depth and Stahel-Donoho outlyingness
   - Skew-adjusted projection depth and adjusted outlyingness
3. Functional data
   - Depth and distance
   - Central tendency and variability of curves
   - Detection of outlying curves
4. Surface and image data

Depth

Univariate data can be ranked!

Depth generalizes rank to other types of data: multivariate observations, regression data, functional data,...

Like rank, depth is a nonparametric notion, since the data are not assumed to come from any distributional model.

Depth function

General idea of a statistical depth function:

Depth function

Given p-variate data $X_n = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ with $\boldsymbol{x}_i \in \mathbb{R}^p$ for all $i = 1, \ldots, n$, a depth function provides an ordering from the outside inward such that the least central objects get the smallest depth values and most central objects get the largest depth.

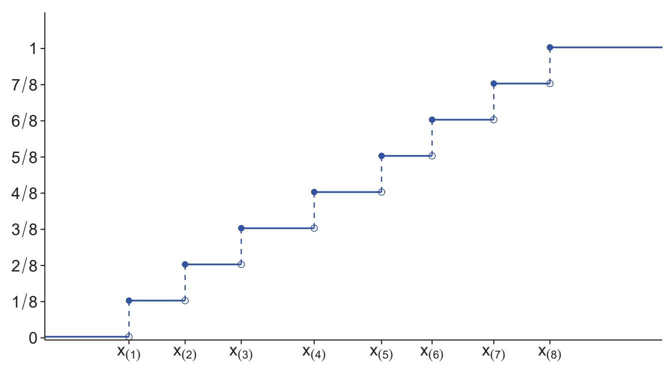
Some desired properties:

1. affine invariant
2. not too sensitive to outliers
3. computationally feasible.

See also the depth axioms in Zuo-Serfling (2000).

Univariate data

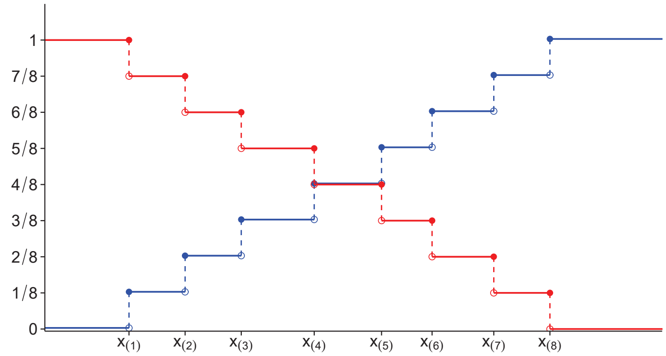
We can rank univariate data $X_n = \{x_1, \ldots, x_n\}$ or equivalently compute the empirical cdf. But the result depends on the orientation of the real line...



$\hat{F}(x; X_n) = \frac{1}{n}\#\{x_i \leqslant x\}$

Univariate data

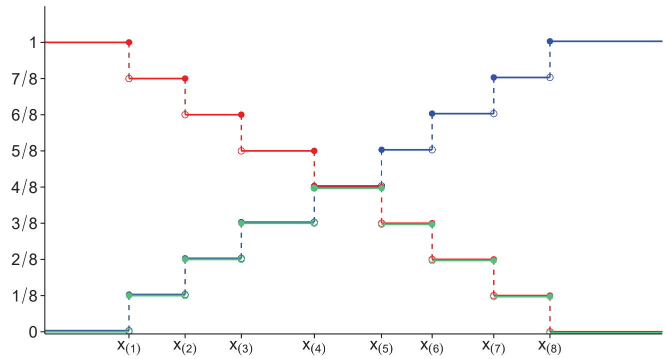
We can rank univariate data $X_n = \{x_1, \ldots, x_n\}$ or equivalently compute the empirical cdf. But the result depends on the orientation of the real line...



$\hat{F}(x; X_n) = \frac{1}{n}\#\{x_i \leqslant x\}$ $\hat{F}(-x; -X_n) = \frac{1}{n}\#\{x_i \geqslant x\}$

Univariate data

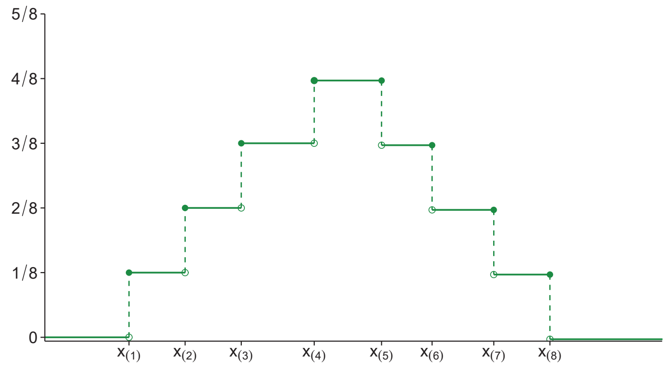
We can rank univariate data $X_n = \{x_1, \ldots, x_n\}$ or equivalently compute the empirical cdf. But the result depends on the orientation of the real line...



$\hat{F}(x; X_n) = \frac{1}{n}\#\{x_i \leqslant x\}$ $\hat{F}(-x; -X_n) = \frac{1}{n}\#\{x_i \geqslant x\}$

$\text{HD}(x; X_n) = \frac{1}{n}\min(\#\{x_i \leqslant x\}, \#\{x_i \geqslant x\})$

Univariate data



$$\begin{aligned}\text{HD}(x; X_n) &= \frac{1}{n}\min(\#\{x_i \leqslant x\}, \#\{x_i \geqslant x\}) \\ &= \frac{1}{n}\min_{u\in\{-1,1\}} \#\{x_i;\ \ ux_i \geqslant ux\}\ .\end{aligned}$$

Univariate data

Formally, for any univariate data set $X_n = \{x_1, \ldots, x_n\}$ and any arbitrary point x we define the depth of x with respect to X_n as:

$$\text{HD}(x; X_n) = \frac{1}{n}\min_{u\in\{-1,1\}} \#\{x_i;\ \ ux_i \geqslant ux\}.$$

Note that x does not have to be an observation!

When n is odd, there is a unique point with largest depth.
When n is even, the largest depth is attained in an interval between two observations.

In general, the depth median is defined as the average of all points with largest depth.

In this setting, the depth median coincides with the usual sample median.

Halfspace depth

Tukey (1975) introduced the concept of halfspace depth to measure the centrality of an arbitrary point within a multivariate data cloud.

Let $\boldsymbol{x} \in \mathbb{R}^p$ be an arbitrary point and consider the data set $X_n = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ with $\boldsymbol{x}_i \in \mathbb{R}^p$ for all $i = 1, \ldots, n$.

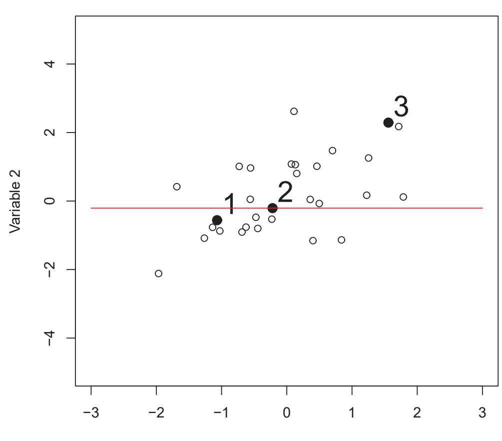
The halfspace depth of $\boldsymbol{x}$ with respect to X_n is defined as the smallest proportion of data points contained in any closed halfspace of which the boundary passes through $\boldsymbol{x}$. More formally:

Halfspace depth

$$\mathsf{HD}(\boldsymbol{x}, X_n) = \frac{1}{n} \min_{\|\boldsymbol{u}\|=1} \#\{\boldsymbol{x}_i;\ \ \boldsymbol{u}'\boldsymbol{x}_i \geqslant \boldsymbol{u}'\boldsymbol{x}\}\ .$$

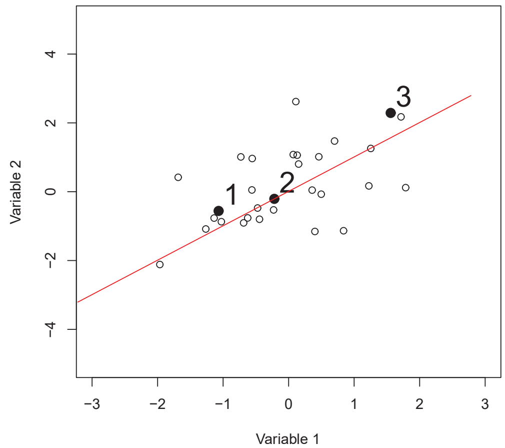
Halfspace depth

Consider the following data set. To compute the halfspace depth of observation 2, we consider all lines through that point and count the minimal number of observations on each side of that line (including points on the line itself).

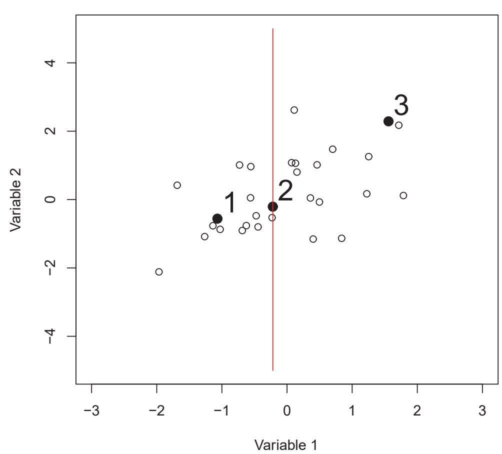


Halfspace depth

Around point 2, every halfplane will contain many observations:

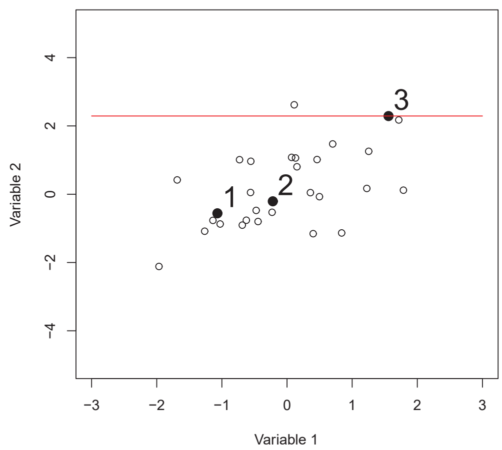


Halfspace depth

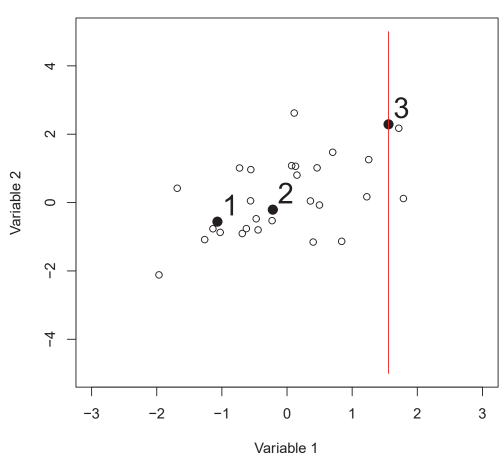


Halfspace depth

When we look at the halfplanes with boundary line through observation 3, some of them contain few observations:

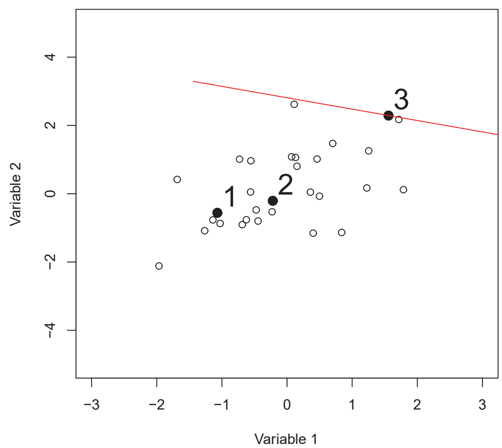


Halfspace depth



Halfspace depth

As observation 3 lies on the convex hull of the data set, its halfspace depth is equal to $1/n$.
All points $\boldsymbol{x}$ outside the convex hull have zero halfspace depth.

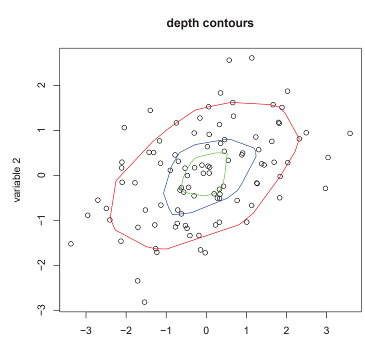


Halfspace depth regions

Depth region

The depth region D_α of level α (with $0 < \alpha < 1$) is the set of points for which $\text{HD}(\boldsymbol{x}, X_n) \geqslant \alpha$.

Halfspace depth regions are bounded convex sets, nested for increasing α.



Computation of halfspace depth

To compute the halfspace depth of a point $\boldsymbol{x}$ in $\mathbb{R}^p$:

- $p = 2, 3$: fast exact computation (Rousseeuw and Ruts 1996; Rousseeuw and Struyf 1998), available in R-package `mrfDepth`.
- $p \leqslant 4, 5$: exact computation (Bremner et al. 2008, Liu and Zuo 2014, Dyckerhoff and Mozharovskyi 2014)
- approximate algorithms: Rousseeuw and Struyf (1998) and Cuesta-Albertos and Nieto-Reyes (2008), available in R-package `mrfDepth`.

Computation of halfspace depth regions

To compute depth regions and their volumes:

- $p = 2$: exact computation of the depth contour and its volume (Ruts and Rousseeuw 1996), available in the R package `mrfDepth`.
- $p \leqslant 5$: exact computation using algorithms in Hallin-Paindaveine-Šiman (2010) and Paindaveine-Šiman (2012).
- approximately: intersections with the depth contours are searched on lines originating from the depth median (bisection algorithm), see `mrfDepth`. Or compute the volume of the convex hull of the data points with depth at least α, using the `convhulln` function in the R-package `geometry`.

Tukey median: definition

The most central point (not necessarily an observation) is the point $\boldsymbol{x}$ with the largest halfspace depth. Often, this point is not unique!

Tukey median

The Tukey median is defined as the average of the set of points with maximal halfspace depth.

The Tukey median is affine equivariant, since halfspace depth is affine invariant.

Efficient algorithms were constructed to compute the Tukey median (Rousseeuw-Ruts 1998; Miller et al. 2003).

Properties

- If X_n is in general position, then

$$\max_{\boldsymbol{x}} \text{HD}(\boldsymbol{x}, X_n) \leqslant \frac{\lceil \frac{n}{2} \rceil}{n} \approx \frac{1}{2} \, .$$

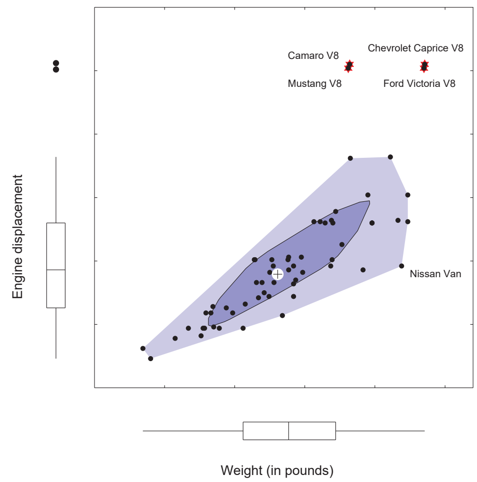
- At any data set X_n it holds that (Rado 1946):

$$\max_{\boldsymbol{x}} \text{HD}(\boldsymbol{x}, X_n) \geqslant \frac{\lceil \frac{n}{p+1} \rceil}{n} \approx \frac{1}{p+1} \, .$$

- The breakdown value of the Tukey median can be as low as $1/(p+1)$, but converges to $1/3$ if the regular data are sampled from a distribution that is symmetric about some central point (Donoho-Gasko 1992).

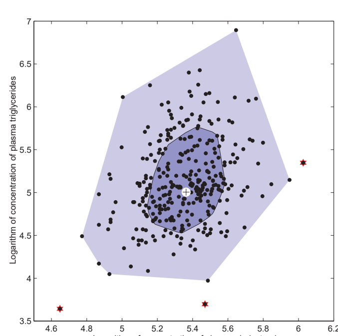
The bagplot

Using halfspace depth we can generalize the well-known univariate boxplot to the bivariate **bagplot** (Rousseeuw-Ruts-Tukey 1999):



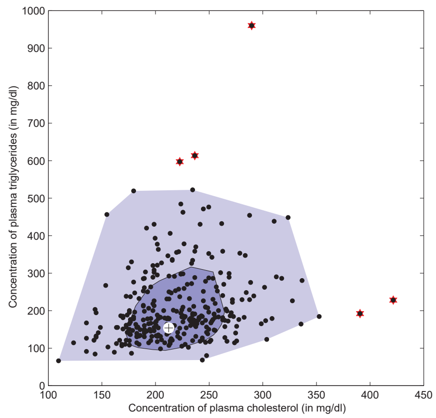
The bagplot: construction

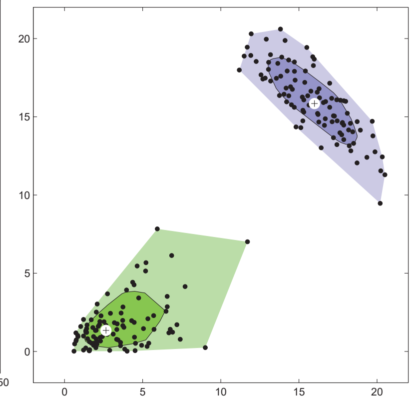
1. The central point [cross] is the Tukey median.
2. The *bag* [dark color] contains 50% of the points.
   (It is interpolated between two depth contours.)
3. The *fence* [not drawn!] inflates the bag about the median by a factor of 3.
4. The *loop* [light color] is the convex hull of all the points inside the fence.
5. The points outside the fence are labeled as outliers [red stars].



The bagplot

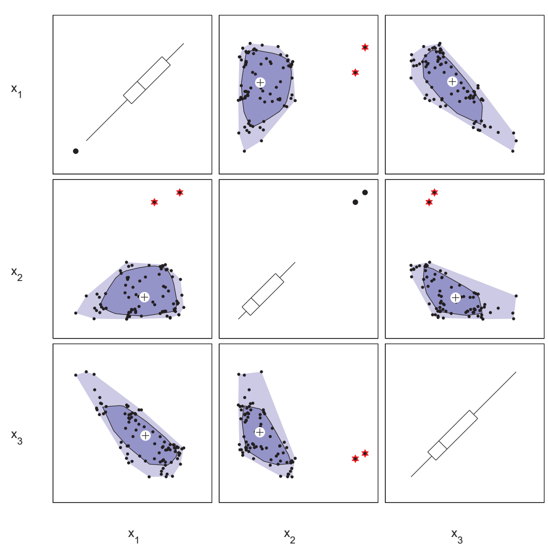
The bagplot visualizes several characteristics of the data: its location, spread (the size of the bag), correlation (the orientation of the bag), skewness (the shape of the bag and the loop), and tails (the outliers).





The bagplot

The bagplot matrix shows several variables at once:



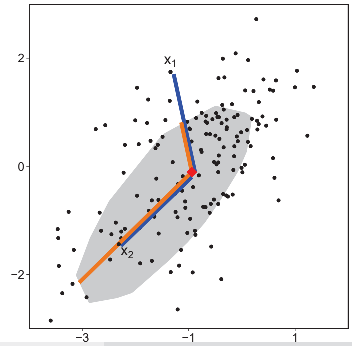
Bagdistance

Bagdistance (Hubert-Rousseeuw-Segaert 2015)

The bagdistance of $\boldsymbol{x} \in \mathbb{R}^p$ to X_n is defined as

$$bd(\boldsymbol{x}; X_n) = \frac{\|\boldsymbol{x} - \boldsymbol{\theta}\|}{\|c_x - \boldsymbol{\theta}\|}$$

- c_x is intersection between the depth contour containing the 50% deepest points (the bag) and the line through $\boldsymbol{x}$ and $\boldsymbol{\theta}$
- Note: outliers on the bagplot have bagdistance $\geqslant 3$. When $p > 2$, $\boldsymbol{x}$ can be flagged as an outlier when $bd(\boldsymbol{x}; X_n) > \sqrt{\chi^2_{p,0.99}}$



Projection depth

Stahel-Donoho outlyingness

The Stahel-Donoho outlyingness of a point $\boldsymbol{x}$ relative to a data set X_n is given by

$$\text{SDO}(\boldsymbol{x}; X_n) = \sup_{\|\boldsymbol{a}\|=1} \frac{|\boldsymbol{a}'\boldsymbol{x} - \text{med}_j(\boldsymbol{a}'\boldsymbol{x}_j)|}{\text{MAD}_j(\boldsymbol{a}'\boldsymbol{x}_j)} .$$

The projection depth (Zuo and Serfling 2000, Zuo 2003) is based on the Stahel-Donoho outlyingness:

Projection depth

The projection depth of a point $\boldsymbol{x}$ relative to a data set X_n is defined as

$$\text{PD}(\boldsymbol{x}; X_n) = \frac{1}{1 + \text{SDO}(\boldsymbol{x}, X_n)} .$$

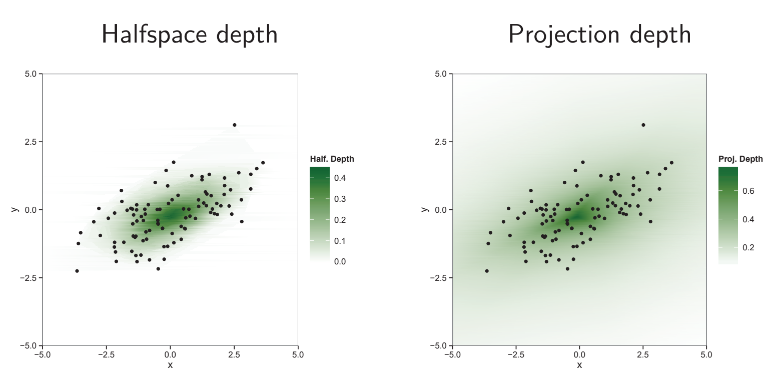
Projection depth

Computation:

- Take the directions orthogonal to hyperplanes spanned by random subsamples of size p. This yields an affine equivariant algorithm (Maronna and Yohai, 1995), available in the R-package `mrfDepth`.
- Exact algorithms available for $p = 2$ (Zuo and Lai, 2011).
- Approximate algorithms (Liu and Zuo, 2014)

Projection depth

Projection depth assigns non-zero values to points outside the convex hull of the data. Hence PD (and SDO) is more appropriate for outlier detection than HD. Also bagdistance gives more information about the degree of outlyingness than HD.



Skew-adjusted projection depth

The adjusted outlyingness of a point $\boldsymbol{x}$ relative to a dataset X_n is defined as (Brys et al. 2005):

$$\mathrm{AO}(\boldsymbol{x}; X_n) = \sup_{\|\boldsymbol{a}\|=1} \mathrm{AO}(\boldsymbol{a}'\boldsymbol{x}; X_n\boldsymbol{a}) \ .$$

The AO is useful for skewed data. Similar to SDO it leads to a depth measure:

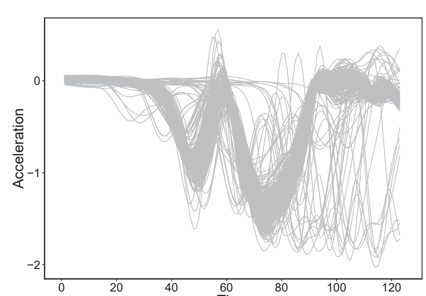
Skew-adjusted projection depth

$$\mathrm{SPD}(\boldsymbol{x}; X_n) = \frac{1}{1 + \mathrm{AO}(\boldsymbol{x}; X_n)} \ .$$

+ Affine invariant
+ Robust
+ Allows for skewness

Univariate functional data

Example: an industrial machine produces one part per cycle. Each cycle is monitored by an accelerometer. A cycle takes 120 ms, measurements are taken every millisecond. The data contains $n = 224$ production cycles.
Several curves have a deviating pattern, most prominently at the final stage of production:



Univariate functional data

Assume that these univariate curves

$$(Y_1(t), t \in U), \ldots, (Y_n(t), t \in U)$$

are realizations of a real-valued stochastic process $\{Y(t), t \in U = [a, b]\}$ with distribution P_Y such that its paths are continuous functions from U to $\mathbb{R}$.

Let us now take an arbitrary continuous function $X : U \to \mathbb{R}$.
How can we define the depth of X relative to the process Y ?

The integrated depth approach (Fraiman and Muniz 2001; Cuevas et al. 2007) starts from a depth function D for univariate data, so at each time point t we can compute $D(X(t); P_Y(t))$. The integrated depth is then defined as

$$\text{ID}(X; P_Y) = \int_U D(X(t); P_Y(t))\, dt\ .$$

Univariate functional data

In the special case where D is the univariate simplicial depth (Liu 1990), the integrated depth becomes the modified band depth MBD (López-Pintado and Romo, 2009).

Other depth measures exist for univariate functional data, such as the random projection depth of (Cuevas et al, 2007).

Note that in practice we don't observe whole curves, but curve evaluations

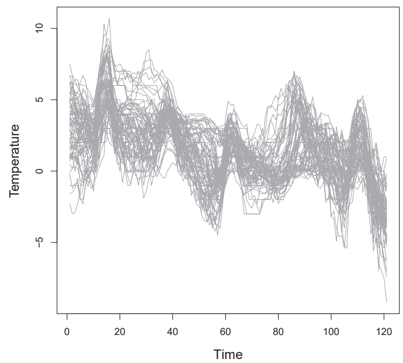
$$\{(Y_1(t_j), \ldots, Y_n(t_j));\ j = 1, \ldots, T\}$$

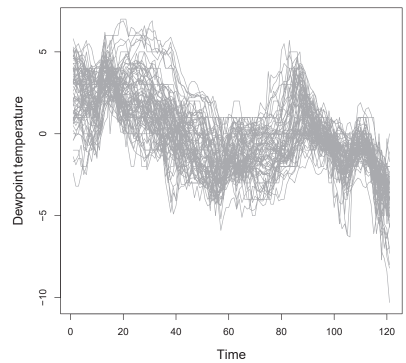
at a set of time points $t_1 < t_2 < \ldots < t_T$ in U (not necessarily equidistant).

Multivariate functional data

Often curves are multivariate in nature!

Example: temperature and dewpoint temperature measured between January 11 and 15, 2013 at 78 weather stations in the U.K.





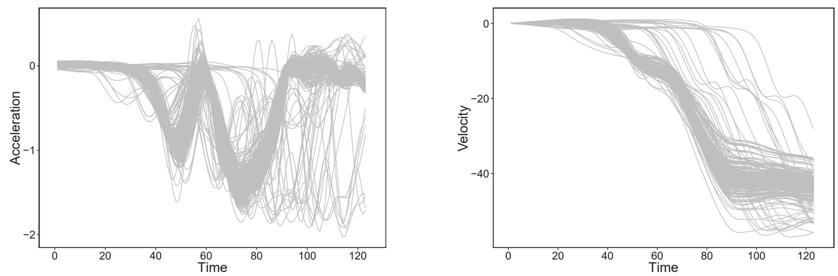
Multivariate functional data

On the other hand, starting from one set of curves one can also compute additional curves such as:

- the first and/or second order **derivatives**. This helps to detect curves with outlying shape.
- the **integrated curves**. When the data consists of acceleration measurements, this helps us to study the velocity as well.
- the **warping functions** that were used to warp the data. This helps us to model phase variations.

Multivariate functional data

Example: acceleration and velocity (integrated!)



Basic questions of interest

1. estimation of the central tendency of the curves
2. estimation of the variability among the curves
3. classification or clustering of such curves
4. detection of outlying curves

Multivariate functional depth

Consider a p-variate stochastic process $\{Y(t), t \in U = [a, b]\}$ with distribution P_Y such that its paths are continuous functions from U to $\mathbb{R}^p$.
Let D be a statistical depth function on $\mathbb{R}^p$ and w a weight function that is defined on U and integrates to one.

Take an arbitrary continuous function $X : U \to \mathbb{R}^p$.

Multivariate functional depth (Claeskens et al. 2014)

The multivariate functional depth (MFD) of X is defined as

$$\text{MFD}(X; P_Y) = \int_U D(X(t); P_Y(t))\, w(t)\, dt\ .$$

When D is the halfspace depth this becomes the Multivariate Functional Halfspace Depth (MFHD).

Weight functions

Examples of weight functions:

- a constant
- a constant times an indicator for a range of interest
- proportional to the volume of the depth region at time point t

$$w(t) = w_\alpha(t; P_Y(t)) = \frac{\text{vol}\{D_\alpha(P_Y(t))\}}{\int_U \text{vol}\{D_\alpha(P_Y(u))\}\, du}$$

- a weight which is large in regions where all curves nearly coincide, in order to penalize outlyingness in these time periods.

When MFHD uses the weight function w_α the resulting depth is denoted by MFHD(α).

MFHD median

Assume $P_Y(t)$ has a unique deepest point for every $t \in U$.

MFHD median

The MFHD median of Y is defined as the curve Θ, in which for each t $\Theta(t)$ is the vector in $\mathbb{R}^p$ with highest value of HD$(\cdot; P_Y(t))$.

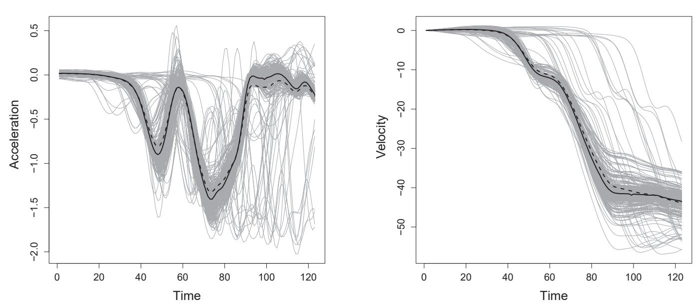
The MFHD median

- does not depend on the weight function
- is affine equivariant (unlike the coordinatewise median)
- is in general not one of the observed curves.
(In the multivariate case it is not one of the data points either.)
- is continuous under several conditions on the depth function and the stochastic process

The median curve of the sample $\{Y_1(t_j), \ldots, Y_n(t_j); j = 1, \ldots, T\}$ is defined as the Tukey median at each time point.

Estimating the central tendency

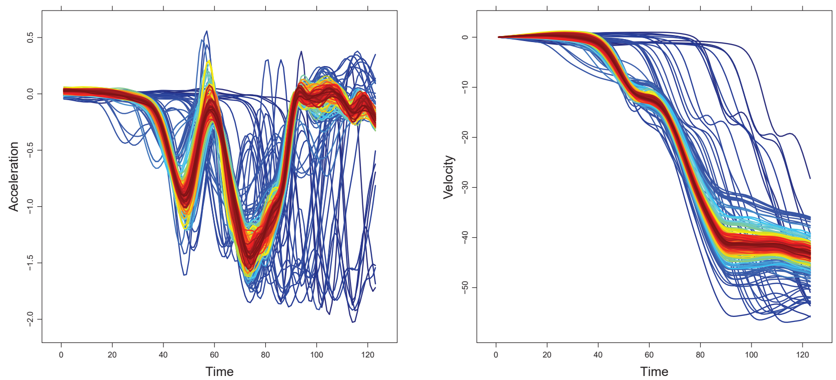
The depth MFHD(α) with $\alpha = 1/8$ yields the MFHD median (solid curve) which is rather robust to the outlying curves.



We can also consider the β-trimmed mean as the mean of the $[n\beta]$ curves with largest depth. This estimator is somewhat more efficient in the gaussian case.

The central tendency and beyond

The rainbow plot (Hyndman and Shang, 2010) colors each curve according to its depth.



Estimating the variability

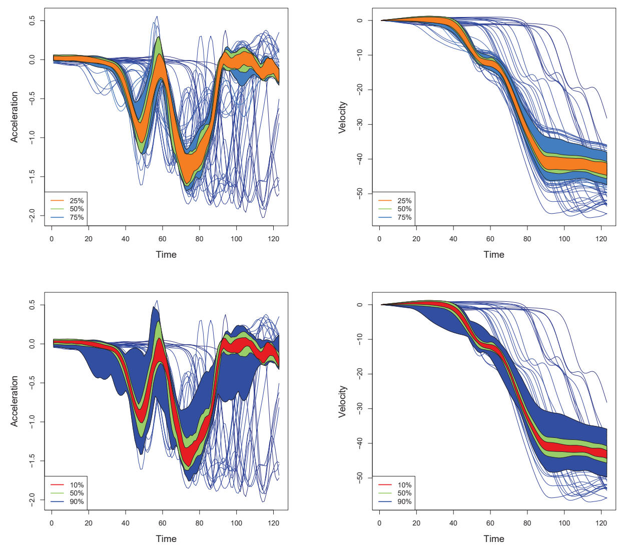
To visualise the variability of the curves, one can plot the central regions (López-Pintado and Romo, 2009), and derive dispersion curves from them.

The β-central region of each component consists of the band delimited by the $[n\beta]$ curves with largest depth.

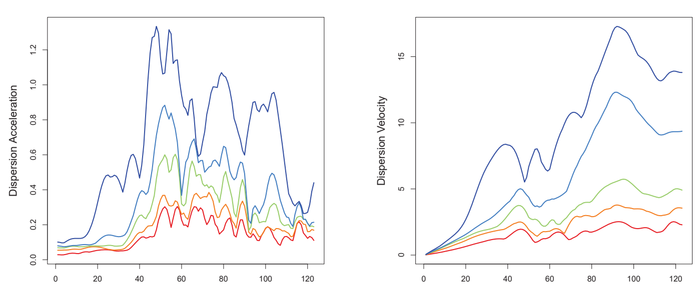
The β-dispersion curve $s_\beta(t)$ of each component is defined as the width of the β-central region at each t.

Typical choices for β: 0.25, 0.5, 0.75. Then for each component $s_{0.5}(t)$ can be seen as a functional IQR (Sun and Genton, 2011).

Central regions



Dispersion curves

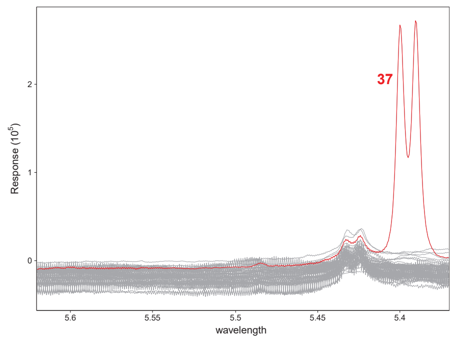


Detecting outlying curves

References: Hubert et al. 2015, Rousseeuw et al. 2016

Isolated outliers: outlying behavior during a short time interval.

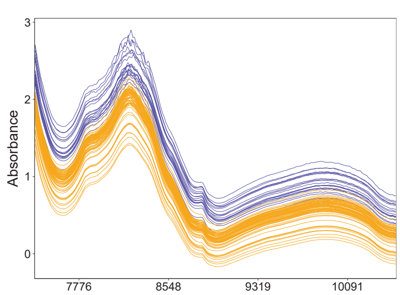
Example: Proton Nuclear Magnetic Resonance spectra of 40 different wine samples.

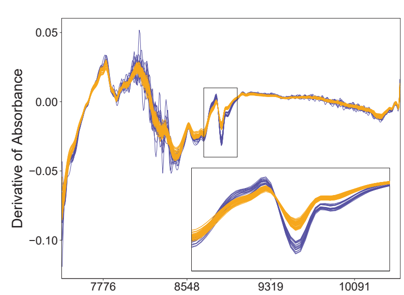


Shape and shift outliers

Shape outliers: shape differs from the majority without necessarily standing out at any time point
Shift outliers: same shape as the majority, but moved away

Example: Near Infrared Spectroscopy responses for a batch of pills. Two groups: 90mg tablets (orange), 250mg (blue).

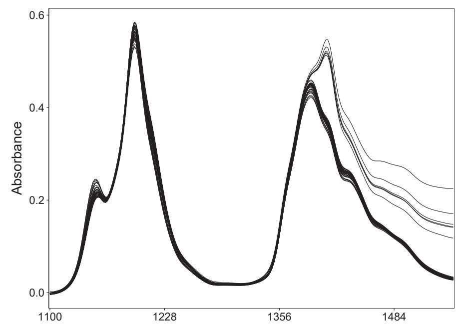




Depth for multivariate functional data

Outlying curves do not always have the lowest depth!

Octane data: 39 Near Infrared spectra of gasoline samples



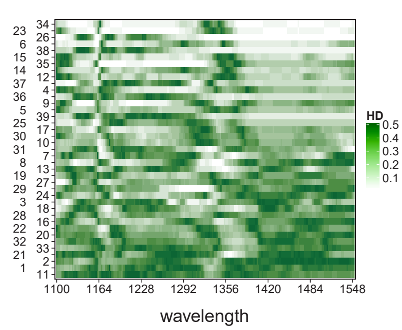
6 samples (25, 26 and 36-39) contain added ethanol, their MFHD ranks are : 16, 3, 12, 10, 5, 15

Depth heatmap

Depth heatmap

- Vertically curves are ordered according to MFD (smallest depth on top).
- Horizontally each entry is colored green according to the multivariate depth at time t.

Octane data:



Distance for multivariate functional data

Functional bagdistance

$$fbd(X;P_Y) = \int_{[a,b]} bd(X(t);P_Y(t)) \cdot \widetilde{w}(t,Y)\,dt$$

Functional adjusted outlyingness

$$f\mathrm{AO}(X;P_Y) = \int_{[a,b]} f\mathrm{AO}(X(t);P_Y(t)) \cdot \widetilde{w}(t,Y)\,dt$$

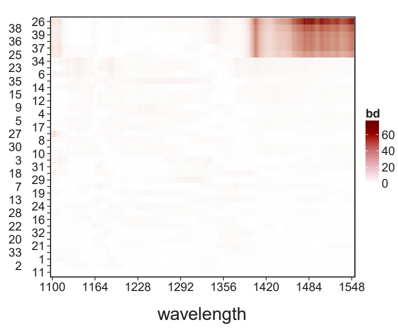
- $\widetilde{w}$ can be different from w
- Under minor conditions fbd and fAO are metrics on set of p-dimensional continuous curves on $[a,b]$.

Distance heatmap

Distance heatmap

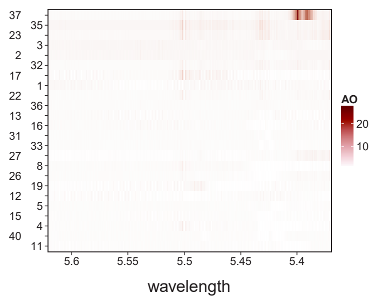
- Vertically curves are ordered according to fbd or fAO (largest distance on top).
- Horizontally each entry is colored red according to the multivariate distance bd or AO at time t.

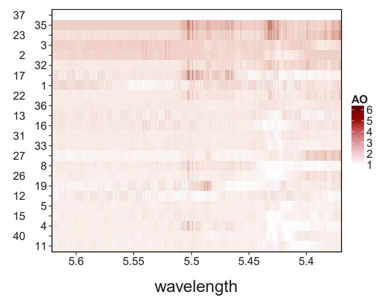
Octane data:



Distance heatmap

Wine data:

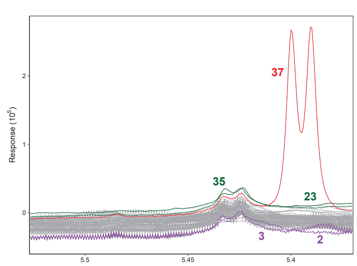


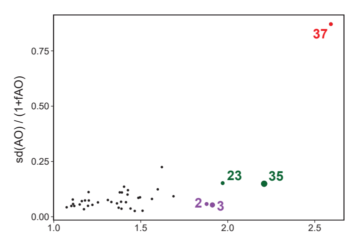


Functional outlier map

Functional outlier map (FOM)

$$\left(\mathsf{fAO}(Y_i;Y)\,,\ \frac{sd_t(\mathsf{AO}(Y_i(t);Y(t)))}{1+\mathsf{fAO}(Y_i;Y)} \right)$$





Shift outliers in the lower right part (since variability in the outlyingness is small), shape outliers in the upper right part.

Flagging outliers

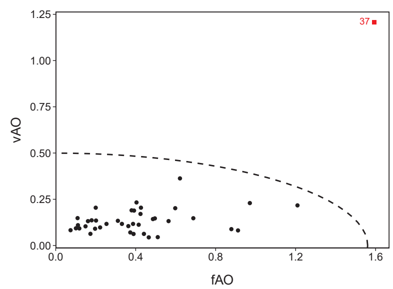
Define the combined functional outlyingness (CFO) of a curve Y_i as

$$\text{CFO}_i = \text{CFO}(Y_i; Y) = \sqrt{(\text{fAO}_i/\text{med}(\text{fAO}))^2 + (\text{vAO}_i/\text{med}(\text{vAO}))^2}$$

where $\text{fAO}_i = \text{fAO}(Y_i; Y)$. We flag a curve Y_i as an outlier iff

$$\frac{\text{LCFO}_i - \text{med}(\text{LCFO})}{\text{MAD}(\text{LCFO})} > \Phi^{-1}(0.995)$$

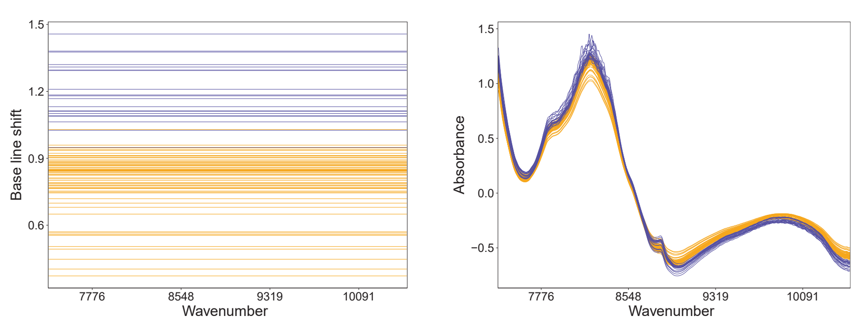
where $\text{LCFO}_i = \log(0.1 + \text{CFO}_i)$.



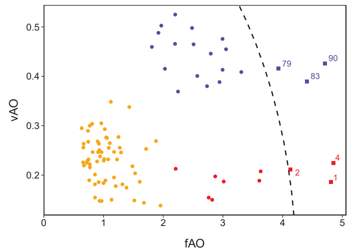
Tablets data: FOM

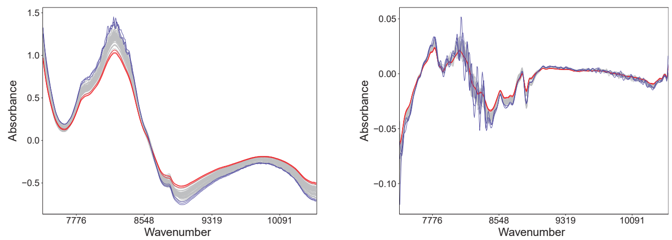
Apply fAO and FOM to:

- mean value of each spectrum (baseline)
- baseline-corrected spectra
- derivatives of the spectra

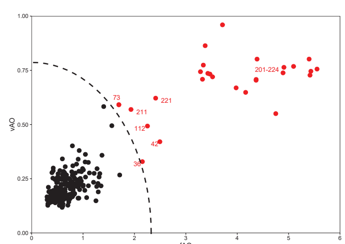


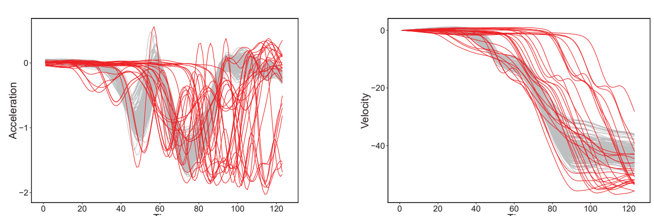
Results





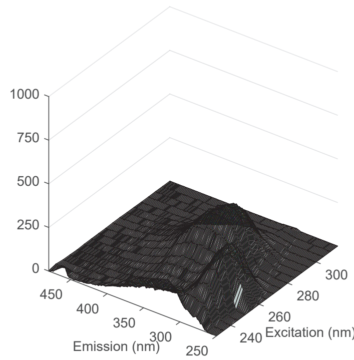
Industrial data: FOM





Dorrit data

Excitation-emission (EEM) landscapes of 27 mixtures of four known fluorophores with excitation wavelengths ranging from 230 nm to 315 nm every 5 nm, and emission at wavelengths from 250 nm to 482 nm at 2 nm intervals.



General framework

Consider a real-valued stochastic process Y with distribution P_Y generating observations:

$$Y_i : [a,b] \times [c,d] \subset \mathbb{R}^2 \to \mathbb{R}^p : (u,v) \mapsto Y_i(u,v) = \begin{pmatrix} Y_i^1(u,v) \\ Y_i^2(u,v) \\ \vdots \\ Y_i^p(u,v) \end{pmatrix}$$

In practice these surfaces are observed on a grid of points, so the i-th observation is actually a $p \times J \times K$ set

$$\{Y_i(j,k);\, j = 1,\ldots,J,\ k = 1,\ldots,K\}$$

A functional data sample of size n is thus a four-dimensional array of size

$$n \times p \times J \times K$$

Functional adjusted outlyingness

Let $Y = \{Y_1, \ldots, Y_n\}$ be a sample of p-variate surfaces recorded at grid points $\{(j,k); j = 1, \ldots, J \text{ and } k = 1, \ldots, K\}$ with $\forall i, j, k : Y_i(j,k) \in \mathbb{R}^p$.

We obtain the adjusted outlyingness of a p-variate surface X with respect to Y by integrating the multivariate adjusted outlyingness over the set of gridpoints.

Functional adjusted outlyingness

$$\text{fAO}(X;Y) = \sum_{j=1}^{J} \sum_{k=1}^{K} \text{AO}(X(j,k); Y(j,k))\, W_{jk}$$

with $\sum_{j=1}^{J} \sum_{k=1}^{K} W_{jk} = 1$.

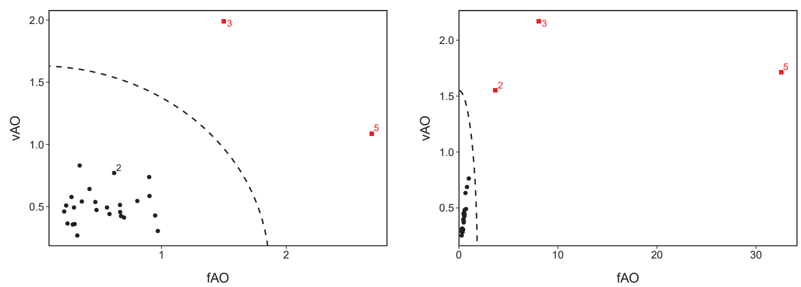
Dorrit data

1. 27 excitation-emission (EEM) landscapes of mixtures of four known fluorophores
2. Y_i contains 18×116 measurements $Y_i(j,k)$ for $j = 1, \ldots, J = 18$ and $k = 1, \ldots, K = 116$.
3. Fit a PARAFAC model:
    1. Trilinear model
    2. Decomposition into the score matrix A and the loading matrices B and C, plus an error term:

    $$Y_i(j,k) = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}\ .$$

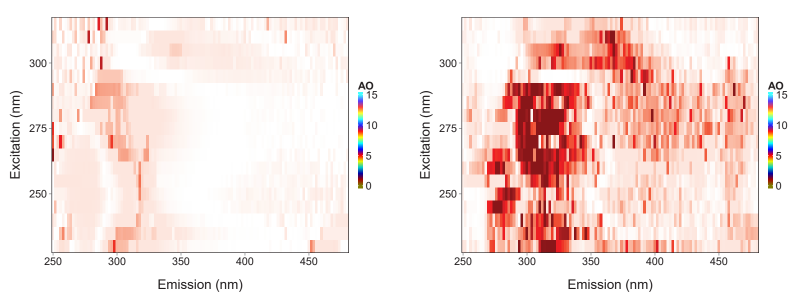
    3. $F = 4$ corresponding with the 4 fluorophores present in the mixtures
4. Continue with the residuals of this model.

Dorrit data: FOMs



Dorrit data: (left) FOM of original data; (right) FOM of residuals after PARAFAC

Dorrit data: heatmaps



Dorrit data, observation 2: (left) AOmap on original data; (right) AOmap on residuals after PARAFAC

References

- Bremner, D., Chen, D., Iacono, J., Langerman, S., Morin, P. (2008). Output-sensitive algorithms for Tukey depth and related problems, *Statistics and Computing*, 18, 259–266.
- Brys, G., Hubert, M., Rousseeuw, P.J. (2005). A robustification of Independent Component Analysis, *Journal of Chemometrics,* 19, 364–375.
- Claeskens, G., Hubert, M., Slaets, L., Vakili, K. (2014). Multivariate functional halfspace depth, *Journal of the American Statistical Association,* 109, 411–423.
- Cuesta-Albertos, J.A., Nieto-Reyes, A. (2008). The random Tukey depth, *Computational Statistics & Data Analysis,* 52, 4979–4988.
- Cuevas, A., Febrero, M., Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions, *Computational Statistics,* 22, 481–496.
- Donoho, D., Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness, *The Annals of Statistics,* 20, 1803–1827.
- Dyckerhoff, R., Mozharovskyi, P. (2014). Exact computation of the halfspace depth, *arXiv:1411.6927*.

References

- Fraiman, R., Muniz, G. (2001). Trimmed means for functional data, *Test,* 10, 419–440.
- Hallin, M., Paindaveine, D., and Šiman, M. (2010). Multivariate quantiles and multiple-output regression quantiles: From L1-optimization to halfspace depth, *The Annals of Statistics*, 38, 635–669.
- Hubert, M., Rousseeuw, P.J., Segaert, P. (2015). Multivariate functional outlier detection (discussion paper with rejoinder), *Statistical Methods & Applications,* 24, 177–202.
- Hyndman, R., Shang, H. (2010). Rainbow plots, bagplots, and boxplots for functional data, *Journal of Computational and Graphical Statistics,* 19, 29–45.
- Liu, X., Zuo, Y. (2014). Computing halfspace depth and regression depth, *Communications in Statistics - Simulation and Computation,* 43, 969–985.
- López-Pintado, S., Romo, J. (2009). On the concept of depth for functional data, *Journal of the American Statistical Association,* 104, 718–734.
- Maronna, R., Yohai, V. (1995). The behavior of the Stahel-Donoho robust multivariate estimator, *Journal of the American Statistical Association,* 90, 330–341.

References

- Miller, K., Ramaswami, S., Rousseeuw, P.J., Sellares, J.A., Souvaine, D., Streinu, I., Struyf, A. (2003). Efficient computation of location depth contours by methods of computational geometry, *Statistics and Computing*, 13, 153–162.
- Paindaveine, D., and Šiman, M. (2012). Computing multiple-output regression quantile regions from projection quantiles, *Computational Statistics*, 27, 29–49.
- Rousseeuw, P.J., Raymaekers, J., Hubert, M. (2016). A measure of directional outlyingness with applications to image data and video, *arXiv:1608.05012*.
- Rousseeuw, P.J., Ruts, I. (1996). Bivariate location depth, *Applied Statistics,* 45, 516–526.
- Rousseeuw, P.J., Ruts, I., Tukey, J. (1999). The bagplot: a bivariate boxplot, *The American Statistician,* 53, 382–387.
- Rousseeuw, P.J., Struyf, A. (1998). Computing location depth and regression depth in higher dimensions, *Statistics and Computing,* 8, 193–203.
- Ruts, I., Rousseeuw, P.J. (1996). Computing depth contours of bivariate point clouds, *Computational Statistics & Data Analysis,* 23, 153–168.

References

- Sun, Y., Genton, M. (2011). Functional boxplots, *Journal of Computational and Graphical Statistics,* 20, 316–334.
- Tukey, J.W. (1975). Mathematics and the Picturing of Data, *Proceedings of the International Congress of Mathematicians,* Vancouver, 2, 523–531.
- Zuo, Y. (2003). Projection-based depth functions and associated medians, *The Annals of Statistics,* 31, 1460–1490.
- Zuo, Y., Lai, S. (2011). Exact computation of bivariate projection depth and the Stahel-Donoho estimator, *Computational Statistics & Data Analysis,* 55, 1173–1179.
- Zuo, Y., Serfling, R. (2000). General notions of statistical depth function, *The Annals of Statistics,* 28, 461–482.