JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2018, VOL. 27, NO. 2, 345–359
https://doi.org/10.1080/10618600.2017.1366912

Taylor & Francis
Taylor & Francis Group

A Measure of Directional Outlyingness With Applications to Image Data and Video

Peter J. Rousseeuw, Jakob Raymaekers, and Mia Hubert

Department of Mathematics, KU Leuven, Leuven, Belgium

ABSTRACT

Functional data analysis covers a wide range of data types. They all have in common that the observed objects are functions of a univariate argument (e.g., time or wavelength) or a multivariate argument (say, a spatial position). These functions take on values which can in turn be univariate (such as the absorbance level) or multivariate (such as the red/green/blue color levels of an image). In practice it is important to be able to detect outliers in such data. For this purpose we introduce a new measure of outlyingness that we compute at each gridpoint of the functions' domain. The proposed *directional outlyingness* (DO) measure accounts for skewness in the data and only requires $\mathcal{O}(n)$ computation time per direction. We derive the influence function of the DO and compute a cutoff for outlier detection. The resulting heatmap and functional outlier map reflect local and global outlyingness of a function. To illustrate the performance of the method on real data it is applied to spectra, MRI images, and video surveillance data.

ARTICLE HISTORY
Received August 2016
Revised June 2017

KEYWORDS
Functional data; Influence function; Outlier detection; Robustness; Skewness

1. Introduction

Functional data analysis (Ramsay and Silverman 2005; Ferraty and Vieu 2006) is a rapidly growing research area. Often the focus is on functions with a univariate domain, such as time series or spectra. The function values may be multivariate, such as temperatures measured at 3, 9, and 12 cm below ground (Berrendero, Justel, and Svarc 2011) or human ECG data measured at eight different places on the body (Pigoli and Sangalli 2012). In this article we will also consider functions whose *domain* is multivariate. In particular, images and surfaces are functions on a bivariate domain. Our methods generalize to higher-dimensional domains, for example the voxels of a three-dimensional image of a human brain are defined on a trivariate domain.

Detecting outliers in functional data is an important task. Recent developments include the approaches of Febrero-Bande, Galeano, and González-Manteiga (2008) and Hyndman and Shang (2010). Sun and Genton (2011) proposed the functional boxplot, and Arribas-Gil et al. (2014) developed the outliergram. Our approach is somewhat different. To detect outlying functions or outlying parts of a function (in a dataset consisting of several functions) we will look at its (possibly multivariate) function value in every time point/pixel/voxel/... of its domain. For this purpose we need a tool that assigns a measure of outlyingness to every data point in a multivariate nonfunctional sample. A popular measure is the Stahel-Donoho outlyingness (SDO) due to Stahel (1981) and Donoho (1982) which works best when the distribution of the inliers is roughly elliptical. However, it is less suited for skewed data. To address this issue, Brys, Hubert, and Rousseeuw (2005) proposed the (skewness-) adjusted outlyingness (AO) which takes the skewness of the underlying distribution into account. However, the AO has two drawbacks. The first is that the AO scale has a large bias as soon as the contamination fraction exceeds 10%. Furthermore, its computation time is $\mathcal{O}(n\log(n))$ per direction due to its rather involved construction.

To remedy these deficiencies we propose a new measure in this article, the *directional outlyingness* (DO). The DO also takes the skewness of the underlying distribution into account, by the intuitive idea of splitting a univariate dataset in two half samples around the median. The AO incorporates a more robust scale estimator, which requires only $\mathcal{O}(n)$ operations.

Section 2 defines the DO, investigates its theoretical properties, and illustrates it on univariate, bivariate, and spectral data. Section 3 derives a cutoff value for the DO and applies it to outlier detection. It also extends the functional outlier map of Hubert, Rousseeuw, and Segaert (2015) to the DO, and in it constructs a curve separating outliers from inliers. Section 4 shows an application to MRI images, and Section 5 analyzes video data. Section 6 contains simulations in various settings, to study the behavior of DO and compare its performance to other methods. Section 7 concludes.

2. A Notion of Directional Outlyingness

2.1. Univariate Setting

In the univariate setting, the Stahel-Donoho outlyingness of a point y relative to a sample $Y = \{y_1, \ldots, y_n\}$ is defined as

$$\mathrm{SDO}(y; Y) = \frac{|y - \mathrm{med}(Y)|}{\mathrm{MAD}(Y)}, \tag{1}$$

where the denominator is the median absolute deviation (MAD) of the sample, given by $\mathrm{MAD}(Y) = \mathrm{med}_i(|y_i - \mathrm{med}_j(y_j)|)/\Phi^{-1}(0.75)$ where Φ is the standard normal cdf. The SDO is affine invariant, meaning that it remains the same

CONTACT Peter J. Rousseeuw peter@rousseeuw.net Department of Mathematics, KU Leuven, Belgium.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/JCGS.

when a constant is added to Y and y, and also when they are multiplied by a nonzero constant.

A limitation of the SDO is that it implicitly assumes the inliers (i.e., the nonoutliers) to be roughly symmetrically distributed. But when the inliers have a skewed distribution, using the MAD as a single measure of scale does not capture the asymmetry. For instance, when the data stem from a right-skewed distribution, the SDO may become large for inliers on the right-hand side, and not large enough for actual outliers on the left-hand side.

This observation led to the (skewness-) adjusted outlyingness (AO) proposed by Brys, Hubert, and Rousseeuw (2005). This notion employs a robust measure of skewness called the medcouple (Brys, Hubert, and Struyf 2004), which however requires $\mathcal{O}(n\log(n))$ computation time. Moreover, we will see in the next subsection that it leads to a rather large explosion bias.

In this article we propose the notion of directional outlyingness (DO) which also takes the potential skewness of the underlying distribution into account, while attaining a smaller computation time and bias. The main idea is to split the sample into two half samples, and then to apply a robust scale estimator to each of them.

More precisely, let $y_1 \leqslant y_2 \leqslant \cdots \leqslant y_n$ be a univariate sample. (The actual algorithm does not require sorting the data.) We then construct two subsamples of size $h = \lfloor \frac{n+1}{2} \rfloor$ as follows. For even n we take $Y_a = \{y_{h+1}, \ldots, y_n\}$ and $Y_b = \{y_1, \ldots, y_h\}$ where the subscripts a and b stand for **a**bove and **b**elow the median. For odd n we put $Y_a = \{y_h, \ldots, y_n\}$ and $Y_b = \{y_1, \ldots, y_h\}$ so that Y_a and Y_b share one data point and have the same size.

Next, we compute a scale estimate for each subsample. Among many available robust estimators we choose a one-step M-estimator with Huber ρ-function due to its fast computation and favorable properties. We first compute initial scale estimates

$$\mathrm{s}_{o,a}(Y) = \mathrm{med}(Z_a)/\Phi^{-1}(0.75) \qquad \text{and}$$
$$\mathrm{s}_{o,b}(Y) = \mathrm{med}(Z_b)/\Phi^{-1}(0.75),$$

where $Z_a = Y_a - \mathrm{med}(Y)$ and $Z_b = \mathrm{med}(Y) - Y_b$ and where $\Phi^{-1}(0.75)$ ensures consistency for Gaussian data. The one-step M-estimates are then given by

$$\mathrm{s}_a(Y) = \mathrm{s}_{o,a}(Y)\sqrt{\frac{1}{2\alpha h}\sum_{z_i \in Z_a} \rho_c\left(\frac{z_i}{\mathrm{s}_{o,a}(Y)}\right)}$$
$$\mathrm{s}_b(Y) = \mathrm{s}_{o,b}(Y)\sqrt{\frac{1}{2\alpha h}\sum_{z_i \in Z_b} \rho_c\left(\frac{z_i}{\mathrm{s}_{o,b}(Y)}\right)}, \qquad (2)$$

where again $h = \lfloor \frac{n+1}{2} \rfloor$ and where $\alpha = \int_0^\infty \rho_c(x)d\Phi(x)$. Here ρ_c denotes the Huber rho function for scale $\rho_c(t) = (\frac{t}{c})^2 \mathbb{1}_{[-c,c]} + \mathbb{1}_{(-\infty,c)\cup(c,\infty)}$ with c a tuning parameter regulating the trade-off between efficiency and bias.

Finally, the DO of a point y relative to a univariate sample $Y = \{y_1, \ldots, y_n\}$ is given by

$$\mathrm{DO}(y; Y) = \begin{cases} \frac{y - \mathrm{med}(Y)}{\mathrm{s}_a(Y)} & \text{if } y \geqslant \mathrm{med}(Y) \\ \frac{\mathrm{med}(Y) - y}{\mathrm{s}_b(Y)} & \text{if } y \leqslant \mathrm{med}(Y). \end{cases} \qquad (3)$$

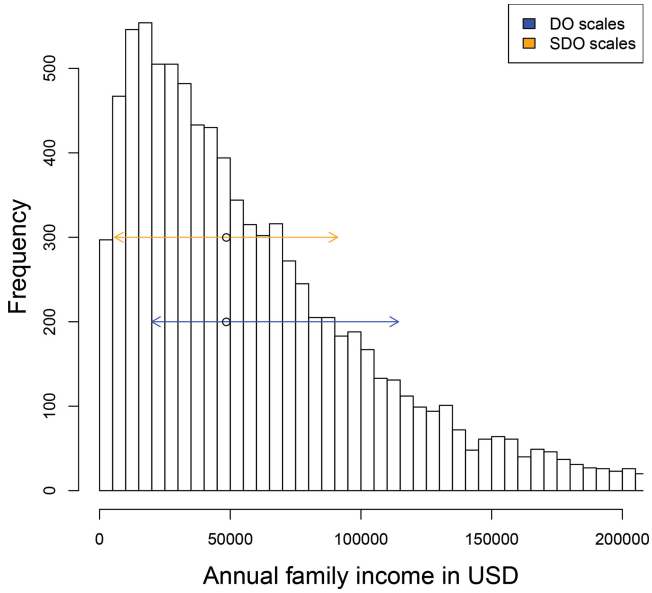

Figure 1. Scale estimates of the family income data. The SDO scale is measured symmetrically about the median, whereas the DO scales are not and reflect skewness.

Note that DO is affine invariant. In particular, flipping the signs of Y and y interchanges s_a and s_b which results in $\mathrm{DO}(-y; -Y) = \mathrm{DO}(y; Y)$.

Figure 1 illustrates the denominators of the SDO and DO expressions on the family income dataset from *https://psidonline.isr.umich.edu* which contains 8962 strictly positive incomes in the tax year 2012. Their histogram is clearly right-skewed. The MAD in the denominator of SDO equals \$42,650 and is used both to the left and to the right of the median, as depicted by the orange arrows. For the DO the "above" scale $\mathrm{s}_a = \$58{,}681$ exceeds the "below" scale $\mathrm{s}_b = \$35{,}737$ (blue arrows). Therefore, a point to the right of the median will have a lower DO than a point to the left at the same distance to the median. This is a desirable property in view of the difference between the left and right tails.

2.2. Robustness Properties

Let us now study the robustness properties of the scales s_a and s_b and the resulting DO. It will be convenient to write s_a and s_b as functionals of the data distribution F:

$$\mathrm{s}_a^2(F) = \frac{\mathrm{s}_{o,a}^2(F)}{\alpha}\int_{\mathrm{med}(F)}^{\infty} \rho_c\left(\frac{x - \mathrm{med}(F)}{\mathrm{s}_{o,a}(F)}\right) dF(x)$$
$$\mathrm{s}_b^2(F) = \frac{\mathrm{s}_{o,b}^2(F)}{\alpha}\int_{-\infty}^{\mathrm{med}(F)} \rho_c\left(\frac{\mathrm{med}(F) - x}{\mathrm{s}_{o,b}(F)}\right) dF(x), \qquad (4)$$

where ρ_c is the Huber ρ-function.

We will first focus on the worst-case bias of s_a due to a fraction ε of contamination, following Martin and Zamar (1993). At a given distribution F, the explosion bias curve of s_a is defined as

$$B^+(\varepsilon, \mathrm{s}_a, F) = \sup_{G \in \mathcal{F}_\varepsilon} (\mathrm{s}_a(G)),$$

where $\mathcal{F}_\varepsilon = \{G : G = (1-\varepsilon)F + \varepsilon H\}$ in which H can be any distribution. The implosion bias curve is defined similarly as

$$B^-(\varepsilon, \mathrm{s}_a, F) = \inf_{G \in \mathcal{F}_\varepsilon} (\mathrm{s}_a(G)).$$

From here onward we will assume that F is symmetric about some center m and has a continuous density $f(x)$ which is strictly decreasing in $x > m$. To derive the explosion and implosion bias we require the following lemma (all proofs can be found in the Appendix):

Lemma 1.

(i) For fixed μ it holds that $t^2 \int_\mu^\infty \rho_c(\frac{x-\mu}{t})\, dF(x)$ is strictly increasing in $t > 0$.

(ii) For fixed $\sigma > 0$ it holds that $\sigma^2 \int_t^\infty \rho_c(\frac{x-t}{\sigma}) dF(x)$ is strictly decreasing in t.

Proposition 1. For any $0 < \varepsilon < 0.25$ the implosion bias of s_a is given by

$$B^-(\varepsilon, \mathrm{s}_a, F)^2 = \frac{1}{\alpha} B^-(\varepsilon, \mathrm{s}_{o,a}, F)^2 \times \left\{ (1-\varepsilon) \int_{B^+(\varepsilon, \mathrm{med}, F)}^\infty \rho_c\left(\frac{x - B^+(\varepsilon, \mathrm{med}, F)}{B^-(\varepsilon, \mathrm{s}_{o,a}, F)} \right) dF(x) \right\},$$

where

$$B^+(\varepsilon, \mathrm{med}, F) = F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$$
$$B^-(\varepsilon, \mathrm{s}_{o,a}, F) = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \left\{ F^{-1}\left(\frac{3-4\varepsilon}{4(1-\varepsilon)}\right) - F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) \right\}.$$

In fact, the implosion bias of s_a is reached when $H = \Delta(F^{-1}(\frac{1}{2(1-\varepsilon)}))$ is the distribution that puts all its mass in the point $F^{-1}(\frac{1}{2(1-\varepsilon)})$. Note that the *implosion breakdown value* of s_a is 25% because for $\varepsilon \to 0.25$ we obtain $\mathrm{s}_a \to 0$.

Proposition 2. For any $0 < \varepsilon < 0.25$ the explosion bias of s_a is given by

$$B^+(\varepsilon, \mathrm{s}_a, F)^2 = \frac{1}{\alpha} B^+(\varepsilon, \mathrm{s}_{o,a}, F)^2 \times \left\{ (1-\varepsilon) \int_{B^+(\varepsilon, \mathrm{med}, F)}^\infty \rho_c\left(\frac{x - B^+(\varepsilon, \mathrm{med}, F)}{B^+(\varepsilon, \mathrm{s}_{o,a}, F)} \right) dF(x) + \varepsilon \right\},$$

where

$$B^+(\varepsilon, \mathrm{s}_{o,a}, F) = \frac{1}{\Phi^{-1}\left(\frac{3}{4}\right)} \left\{ F^{-1}\left(\frac{3}{4(1-\varepsilon)}\right) - F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) \right\}.$$

The explosion bias of s_a is reached at all distributions $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta(d)$ for which $d > B^+(\varepsilon, \mathrm{med}, F) + cB^+(\varepsilon, \mathrm{s}_{o,a}, F)$ which ensures that d lands on the constant part of ρ_c. For $\varepsilon \to 0.25$ we find $d \to \infty$ and $\mathrm{s}_a \to \infty$, so the *explosion breakdown value* of s_a is 25%.

The blue curves in Figure 2 are the explosion and implosion bias curves of s_a when $F = \Phi$ is the standard Gaussian distribution, and the tuning constant in ρ_c is $c = 2.1$ corresponding to 85% efficiency. By affine equivariance the curves for s_b are exactly the same, so these are the curves of both DO scales. The orange curves correspond to explosion and implosion of the scales used in the adjusted outlyingness AO under the same contamination. We see that the AO scale explodes faster, due to using the medcouple in its definition. The fact that the DO scale is typically smaller enables the DO to attain larger values in outliers.

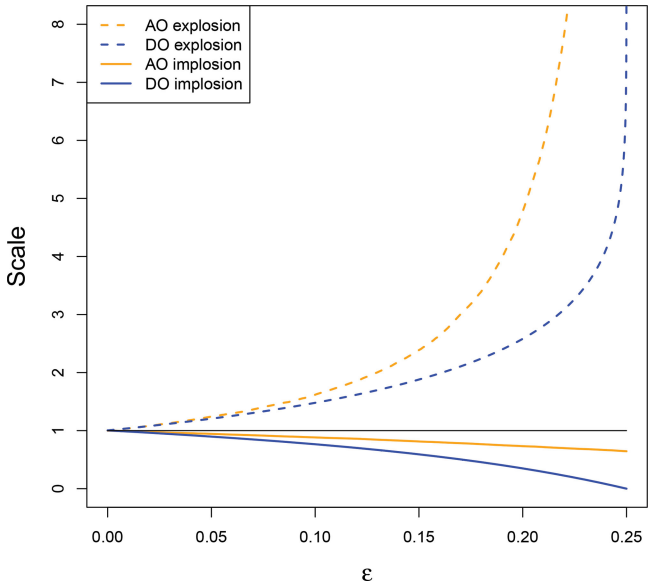

Figure 2. Comparison of explosion and implosion bias of the AO and DO scales.

Another tool to measure the (non-)robustness of a procedure is the influence function (IF). Let T be a statistical functional, and consider the contaminated distribution $F_{\varepsilon,z} = (1-\varepsilon)F + \varepsilon\Delta(z)$. The influence function of T at F is then given by

$$\mathrm{IF}(z, T, F) = \lim_{\varepsilon \to 0} \frac{T(F_{\varepsilon,z}) - T(F)}{\varepsilon} = \frac{\partial}{\partial \varepsilon} T(F_{\varepsilon,z}) \Big|_{\varepsilon=0}$$

and basically describes how T reacts to a small amount of contamination.

This concept justifies our choice for the function ρ_c. Indeed, the IF of a fully iterated M-estimator of scale with function ρ is proportional to $\rho(z) - \beta$ with the constant $\beta = \int_{-\infty}^\infty \rho(x) dF(x)$. We use $\rho = \rho_c$ with $c = 2.1$. It was shown in Hampel et al. (1986) that at $F = \Phi$ this ρ_c yields the M-estimator with highest asymptotic efficiency subject to an upper bound on the absolute value of its IF.

We will now derive the IF of the one-step M-estimator s_a given by (4).

Proposition 3. The influence function of s_a is given by

$$\begin{aligned} 2\alpha \frac{\mathrm{s}_a(F)}{\mathrm{s}_{o,a}^2(F)} \mathrm{IF}(z, \mathrm{s}_a, F) &= \left\{ \frac{2}{\mathrm{s}_{o,a}(F)} \int_{\mathrm{med}(F)}^\infty \rho_c\left(\frac{x - \mathrm{med}(F)}{\mathrm{s}_{o,a}(F)} \right) dF(x) \right. \\ &\quad - \int_{\mathrm{med}(F)}^\infty x \rho_c'\left(\frac{x - \mathrm{med}(F)}{\mathrm{s}_{o,a}(F)} \right) dF(x) \\ &\quad \left. + \mathrm{med}(F) \int_{\mathrm{med}(F)}^\infty \rho_c'\left(\frac{x - \mathrm{med}(F)}{\mathrm{s}_{o,a}(F)} \right) dF(x) \right\} \mathrm{IF}(z, \mathrm{s}_{o,a}, F) \\ &\quad - \left\{ \int_{\mathrm{med}(F)}^\infty \rho_c'\left(\frac{x - \mathrm{med}(F)}{\mathrm{s}_{o,a}(F)} \right) dF(x) \right\} \mathrm{IF}(z, \mathrm{med}, F) \\ &\quad + \left\{ \mathbb{1}_{[\mathrm{med}(F), \infty)}(z) \rho_c(z - \mathrm{med}(F)) - \alpha \right\}, \end{aligned}$$

where $\mathrm{IF}(z, \mathrm{s}_{o,a}, F)$ is the influence function of $\mathrm{s}_{o,a}$.

The resulting IF of s_a at $F = \Phi$ is shown in Figure 3. [Note that $\mathrm{IF}(z, \mathrm{s}_b, \Phi) = \mathrm{IF}(-z, \mathrm{s}_a, \Phi)$.] It is bounded, indicating that s_a is robust to a small amount of contamination even

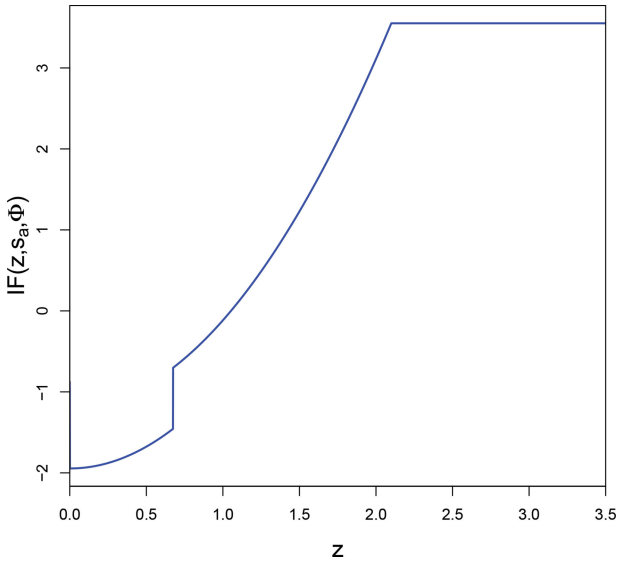

Figure 3. Influence function of s_a at $F = \Phi$.

when it is far away. Note that the IF has a jump at the third quartile $Q_3 \approx 0.674$ due to the initial estimate $s_{o,a}$. If we were to iterate (4) to convergence this jump would vanish, but then the explosion bias would go up a lot, similarly to the computation in Rousseeuw and Croux (1994).

Let us now compute the influence function of $\mathrm{DO}(x; F)$ given by (3) for contamination in the point z, noting that x and z need not be the same.

Proposition 4. When $x > \mathrm{med}(F)$ it holds that

$$\mathrm{IF}(z, DO(x), F) = \frac{-1}{s_a^2(F)}\{\mathrm{IF}(z, \mathrm{med}, F)s_a(F) + \mathrm{IF}(z, s_a, F)(x - \mathrm{med}(F))\}$$

whereas for $x < \mathrm{med}(F)$ we obtain

$$\mathrm{IF}(z, DO(x), F) = \frac{1}{s_b^2(F)}\{\mathrm{IF}(z, \mathrm{med}, F)s_b(F) - \mathrm{IF}(z, s_b, F)(\mathrm{med}(F) - x)\}.$$

For a fixed value of x the influence function of $\mathrm{DO}(x)$ is bounded in z. This is a desirable robustness property. Figure 4 shows the influence function (which is a surface) when F is the standard Gaussian distribution.

2.3. Multivariate Setting

In the multivariate setting we can apply the principle that a point is outlying with respect to a dataset if it stands out in at least one direction. Like the Stahel-Donoho outlyingness, the multivariate DO is defined by means of univariate projections. To be precise, the DO of a d-variate point $\boldsymbol{y}$ relative to a d-variate sample $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ is defined as

$$\mathrm{DO}(\boldsymbol{y}; \boldsymbol{Y}) = \sup_{\boldsymbol{v} \in \mathbb{R}^d} \mathrm{DO}(\boldsymbol{y}^T\boldsymbol{v}; \boldsymbol{Y}^T\boldsymbol{v}), \tag{5}$$

where the right-hand side uses the univariate DO of (3).

To compute the multivariate DO we have to rely on approximate algorithms, as it is impossible to project on *all* directions $\boldsymbol{v}$ in d-dimensional space. A popular procedure to generate a direction is to randomly draw d data points, compute the hyperplane passing through them, and then to take the direction $\boldsymbol{v}$ orthogonal to it. This guarantees that the multivariate DO is affine invariant. That is, the DO does not change when we add a constant vector to the data, or multiply the data by a nonsingular $d \times d$ matrix.

As an illustration we take the bloodfat dataset, which contains plasma cholesterol and plasma triglyceride concentrations (in mg/dl) of 320 male subjects for whom there is evidence of narrowing arteries (Hand et al. 1994). Here $n = 320$ and $d = 2$, and following Hubert and Van der Veeken (2008) we generated $250d = 500$ directions $\boldsymbol{v}$. Figure 5 shows the contour plots of both the DO and SDO measures. Their contours are always convex. We see that the contours of the DO capture the skewness in the dataset, whereas those of the SDO are more symmetric even though the data themselves are not.

2.4. Functional Directional Outlyingness

We now extend the DO to data where the objects are functions. To fix ideas we will consider an example. The glass dataset consists of spectra with $d = 750$ wavelengths resulting from spectroscopy on $n = 180$ archeological glass samples (Lemberge et al. 2000). Figure 6 shows the 180 curves.

At each wavelength the response is a single number, the intensity, so this is a univariate functional dataset. However, we can incorporate the dynamic behavior of these curves by numerically computing their first derivative. This yields

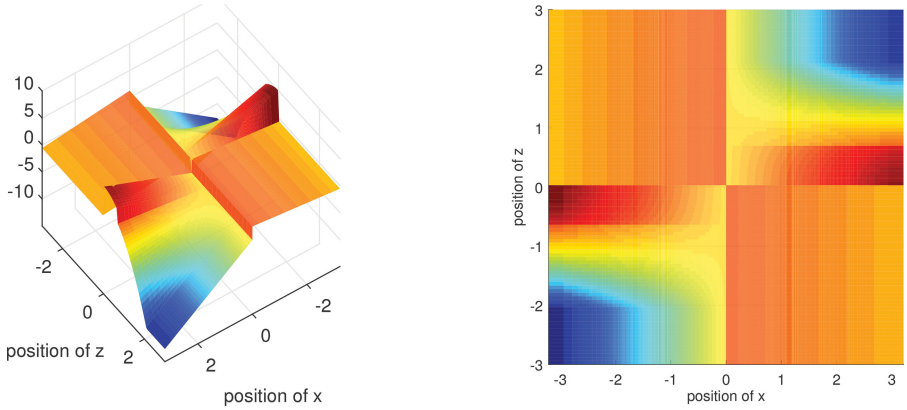

Figure 4. Influence function of $\mathrm{DO}(x)$ for $F = \Phi$. Left: 3D, right: 2D seen from above. For a fixed point x it is bounded over all possible positions z of contamination.

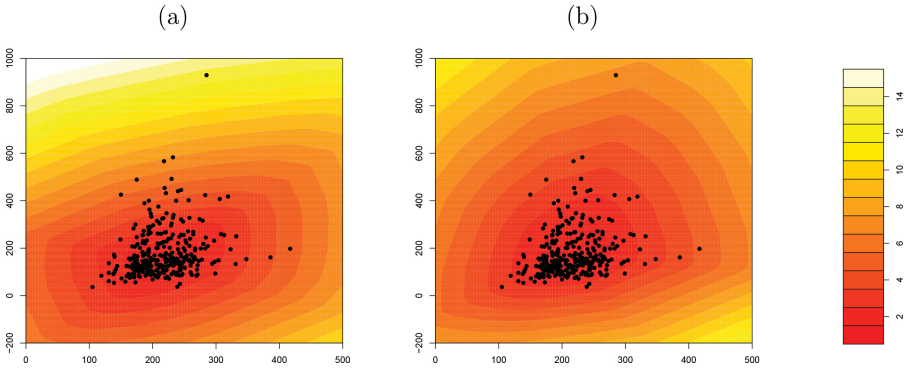

Figure 5. Bloodfat data with (a) SDO contours, and (b) DO contours. The DO contours better reflect the skewness in the data.

bivariate functions, where the response consists of both the intensity and its derivative.

In general we write a functional dataset as $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ where each Y_i is a d-dimensional function. As in this example, the Y_i are typically observed on a discrete set of points in their domain. For a univariate domain this set is denoted as $\{t_1, \ldots, t_T\}$.

Now we want to define the DO of a d-variate function X on the same domain, where X need not be one of the Y_i. For this we look at all the domain points t_j in turn, and define the *functional directional outlyingness* (fDO) of X with respect to the sample $\boldsymbol{Y}$ as

$$\text{fDO}(X; \boldsymbol{Y}) = \sum_{j=1}^{T} \text{DO}(X(t_j); \boldsymbol{Y}(t_j))W(t_j), \tag{6}$$

where $W(.)$ is a weight function for which $\sum_{j=1}^{T} W(t_j) = 1$. Such a weight function allows us to assign a different importance to the outlyingness of a curve at different domain points. For instance, one could downweight time points near the boundaries if measurements are recorded less accurately at the beginning and the end of the process.

Figure 7 shows the fDO of the 180 bivariate functions in the glass data, where $W(.)$ was set to zero for the first 13 wavelengths where the spectra had no variability, and constant at the remaining wavelengths. These fDO values allow us to rank the spectra from most to least outlying, but do not contain much information about how the outlying curves are different from the majority.

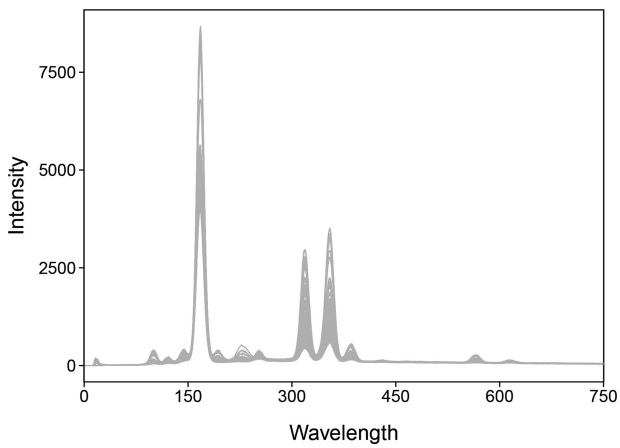

Figure 6. Spectra of 180 archeological glass samples.

In addition to this global outlyingness measure fDO we also want to look at the local outlyingness. To this end Figure 8 shows the individual $\text{DO}(Y_i(t_j); \boldsymbol{Y}(t_j))$ for each of the 180 functions Y_i of the glass data at each wavelength t_j. Higher values of DO are shown by darker red in this heatmap. Now we see that there are a few groups of curves with particular anomalies: one group around function 25, one around function 60, and one with functions near the bottom. Note that the global outlyingness measure fDO flags outlying rows in this heatmap, whereas the dark spots inside the heatmap can be seen as outlying cells. It is also possible to sort the rows of the heatmap according to their fDO values. Note that the wavelength at which a dark spot in the heatmap occurs allows to identify the chemical element responsible.

As in Hubert, Rousseeuw, and Segaert (2015) we can transform the DO to the multivariate depth function $1/(\text{DO} + 1)$, and the fDO to the functional depth function $1/(\text{fDO} + 1)$.

3. Outlier Detection

3.1. A Cutoff for Directional Outlyingness

When analyzing a real dataset we do not know its underlying distribution, but still we would like a rough indication of which observations should be flagged as outliers. For this purpose we need an approximate cutoff value on the DO. We first consider

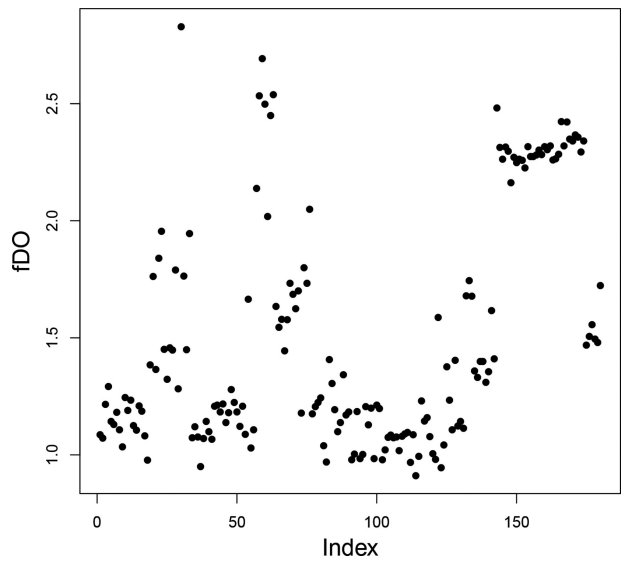

Figure 7. Functional DO (fDO) values of the 180 glass spectra. Higher fDO values correspond to curves that are more outlying on average.

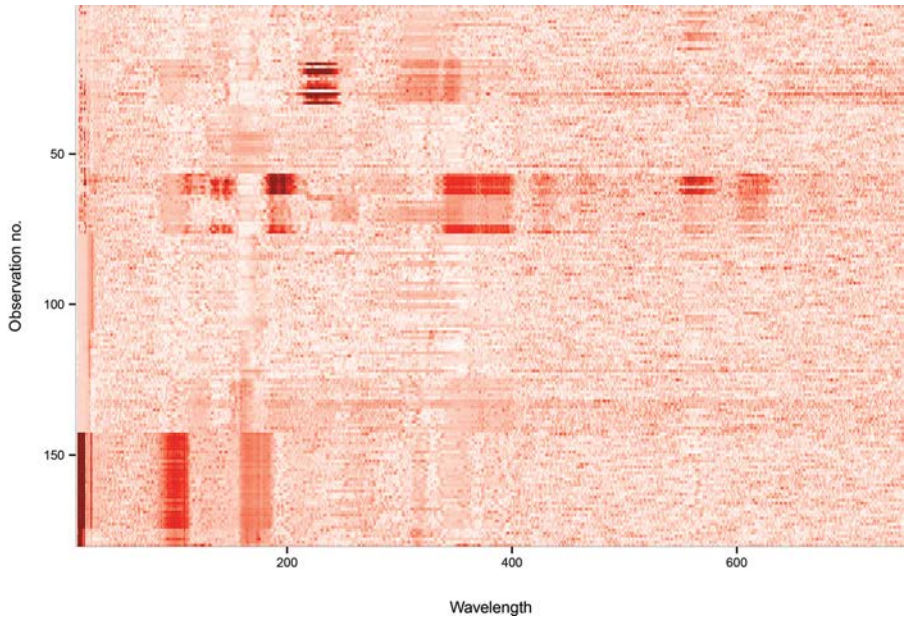

Figure 8. Heatmap of DO of the glass data. Darker pixels indicate outlying behavior.

non-functional data, leaving the functional case for the next subsection. Let $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ be a d-variate dataset ($d \geqslant 1$) with directional outlyingness values $\{\mathrm{DO}_1, \ldots, \mathrm{DO}_n\}$. The DO_i have a right-skewed distribution, so we transform them to $\{\mathrm{LDO}_1, \ldots, \mathrm{LDO}_n\} = \{\log(0.1 + \mathrm{DO}_1), \ldots, \log(0.1 + \mathrm{DO}_n)\}$ of which the majority is closer to Gaussian. Then we center and normalize the resulting values in a robust way and compare them to a high gaussian quantile. For instance, we can flag $\boldsymbol{y}_i$ as outlying whenever

$$\frac{\mathrm{LDO}_i - \mathrm{med}(\mathrm{LDO})}{\mathrm{MAD}(\mathrm{LDO})} > \Phi^{-1}(0.995)\,, \tag{7}$$

so the cutoff for the DO values is $c = \exp\big(\mathrm{med}(LDO) + \mathrm{MAD}(LDO)\Phi^{-1}(0.995)\big) - 0.1$. (Note that we can use the same formulas for functional data by replacing DO by fDO.)

For an illustration we return to the family income data of Figure 1. The blue vertical line in Figure 9 corresponds to the DO cutoff, whereas the orange line is the result of the same computation applied to the SDO. The DO cutoff is the more conservative one, because it takes the skewness of the income distribution into account.

Figure 10 shows the DO and SDO cutoffs for the bivariate bloodfat data of Figure 5. The DO captures the skewness in the data and flags only two points as outlying, whereas the SDO takes a more symmetric view and also flags five of the presumed inliers.

3.2. The Functional Outlier Map

When the dataset consists of functions there can be several types of outlyingness. As an aid to distinguish between them, Hubert, Rousseeuw, and Segaert (2015) introduced a graphical tool called the *functional outlier map* (FOM). Here we will extend the FOM to the new DO measure and add a cutoff to it, to increase its utility.

Consider a functional dataset $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$. The fDO [see (6)] of a function Y_i can be interpreted as the "average outlyingness" of its (possibly multivariate) function values. We now also measure the *variability* of its DO values, by

$$\mathrm{vDO}(Y_i; \boldsymbol{Y}) = \frac{\mathrm{stdev}_j\big(\mathrm{DO}(Y_i(t_j); \boldsymbol{Y}(t_j))\big)}{1 + \mathrm{fDO}(Y_i; \boldsymbol{Y})}. \tag{8}$$

Note that (8) has the fDO in the denominator to measure relative instead of absolute variability. This can be understood as follows. Suppose that the functions Y_i are centered around zero and that $Y_k(t_j) = 2Y_i(t_j)$ for all j. Then $\mathrm{stdev}_j(\mathrm{DO}(Y_k(t_j); \boldsymbol{Y}(t_j))) = 2\mathrm{stdev}_j(\mathrm{DO}(Y_i(t_j); \boldsymbol{Y}(t_j)))$ but their relative variability is the same. Because $\mathrm{fDO}(Y_k; \boldsymbol{Y}) = 2\mathrm{fDO}(Y_i; \boldsymbol{Y})$, putting fDO in the denominator normalizes for this. In the numerator we could also compute a weighted standard deviation with the weights $W(t_j)$ from (6).

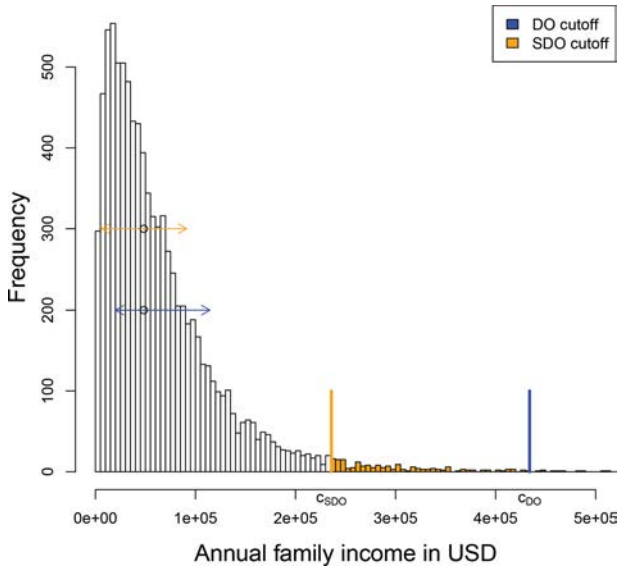

Figure 9. Outlier cutoffs for the family income data. The DO-based cutoff takes the data skewness into account.

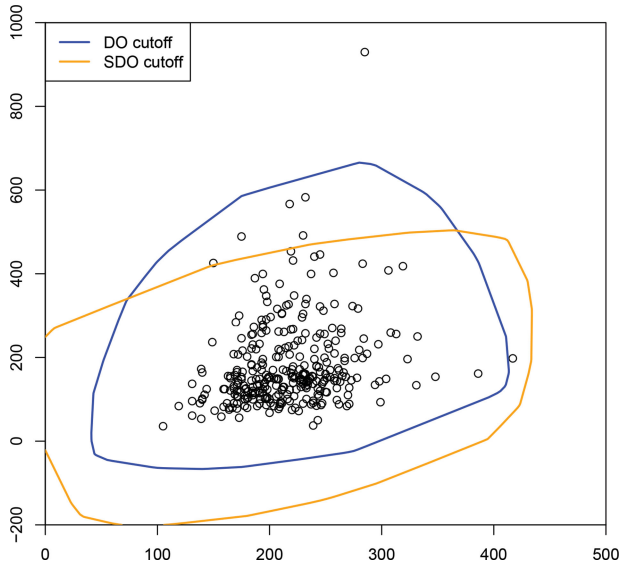

Figure 10. Outlier detection on bloodfat data. The DO-based cutoff adapts to the data skewness and flags fewer points as outlying.

The FOM is then the scatterplot of the points

$$(\mathrm{fDO}(Y_i; \boldsymbol{Y}), \mathrm{vDO}(Y_i; \boldsymbol{Y})) \tag{9}$$

for $i = 1, \ldots, n$. Its goal is to reveal outliers in the data, and its interpretation is fairly straightforward. Points in the lower left part of the FOM represent regular functions which hold a central position in the dataset. Points in the lower right part are functions with a high fDO but a low variability of DO values. This happens for *shift outliers*, that is, functions which have a regular shape but are shifted on the whole domain. Points in the upper left part have a low fDO but a high vDO. Typical examples are local outliers, that is functions which only display outlyingness over a small part of their domain. The points in the upper right part of the FOM have both a high fDO and a high vDO. These correspond to functions which are strongly outlying on a substantial part of their domain.

As an illustration we revisit the glass data. Their FOM in Figure 11 contains a lot more information than their fDO values alone in Figure 7. In the heatmap (Figure 8) we noticed three groups of outliers, which also stand out in the FOM. The first group consists of the spectra 20, 22, 23, 28, 30, 31, and 33. Among these, number 30 lies furthest to the right in the FOM. It corresponds to row 30 in Figure 8 which has a dark red piece. It does not look like a shift outlier, for which the row would have a more homogeneous color (hence a lower vDO). The second group, with functions 57–63, occupies a similar position in the FOM. The group standing out the most consists of functions 143–174. They are situated in the upper part of the FOM, indicating that they are shape outliers. Indeed, they deviate strongly from the majority in three fairly small series of wavelengths. Their outlyingness is thus more local than that of functions 57–63.

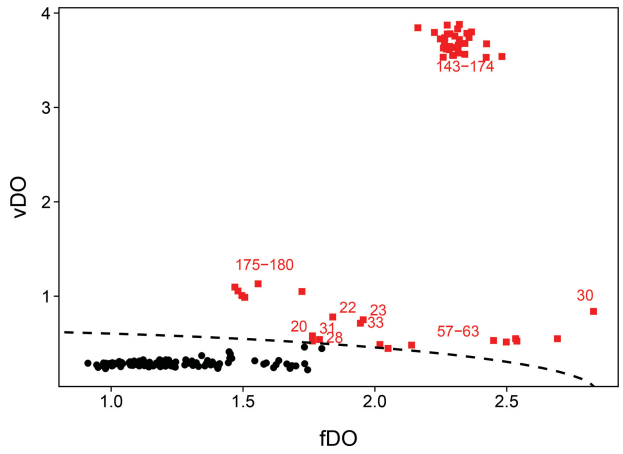

Figure 11. Functional outlier map (FOM) of the glass data, with cutoff curve.

We now add a new feature to the FOM, namely a rule to flag outliers. For this we define the *combined functional outlyingness* (CFO) of a function Y_i as

$$\begin{aligned}\mathrm{CFO}_i &= \mathrm{CFO}(Y_i; \boldsymbol{Y}) \\ &= \sqrt{(\mathrm{fDO}_i/\,\mathrm{med}(\mathrm{fDO}))^2 + (\mathrm{vDO}_i/\,\mathrm{med}(\mathrm{vDO}))^2},\end{aligned} \tag{10}$$

where $\mathrm{fDO}_i = \mathrm{fDO}(Y_i; \boldsymbol{Y})$ and $\mathrm{med}(\mathrm{fDO}) = \mathrm{med}(\mathrm{fDO}_1, \ldots, \mathrm{fDO}_n)$, and similarly for vDO. Note that the CFO characterizes the points in the FOM through their Euclidean distance to the origin, after scaling. We expect outliers to have a large CFO. In general, the distribution of the CFO is unknown but skewed to the right. To construct a cutoff for CFO we use the same reasoning as for the cutoff (7) on fDO: First we compute $\mathrm{LCFO}_i = \log(0.1 + \mathrm{CFO}_i)$ for all $i = 1, \ldots, n$, and then we flag function Y_i as outlying if

$$\frac{\mathrm{LCFO}_i - \mathrm{med}(\mathrm{LCFO})}{\mathrm{MAD}(\mathrm{LCFO})} > \Phi^{-1}(0.995)\,. \tag{11}$$

This yields the dashed curve (which is part of an ellipse) in the FOM of Figure 11.

4. Application to Image Data

Images are functions on a bivariate domain. In practice the domain is a grid of discrete points, for example the horizontal and vertical pixels of an image. It is convenient to use two indices $j = 1, \ldots, J$ and $k = 1, \ldots, K$, one for each dimension of the grid, to characterize these points. An image (or a surface) is then a function on the $J \times K$ points of the grid. Note that the function values can be univariate, like gray intensities, but they can also be multivariate, for example the intensities of red, green, and blue (RGB). In general we will write an image dataset as a sample $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ where each Y_i is a function from $\{(j, k);\ j = 1, \ldots, J \text{ and } k = 1, \ldots, K\}$ to $\mathbb{R}^d$.

The fDO (6) and vDO (8) notions that we saw for functional data with a univariate domain can easily be extended to functions with a bivariate domain by computing

$$\mathrm{fDO}(Y_i; \boldsymbol{Y}) = \sum_{j=1}^{J} \sum_{k=1}^{K} \mathrm{DO}(Y_i(j, k); \boldsymbol{Y}(j, k)) W_{jk}, \tag{12}$$

where the weights W_{jk} must satisfy $\sum_{j=1}^{J} \sum_{k=1}^{K} W_{jk} = 1$, and

$$\mathrm{vDO}(Y_i; \boldsymbol{Y}) = \frac{\mathrm{stdev}_{j,k}(\mathrm{DO}(Y_i(j, k); \boldsymbol{Y}(j, k)))}{1 + \mathrm{fDO}(Y_i; \boldsymbol{Y})}, \tag{13}$$

Figure 12. Original MRI image of subject 387, and its derivatives in the horizontal and vertical direction.

where the standard deviation can also be weighted by the W_{jk}. (The simplest weight function is the constant $W_{jk} = 1/(JK)$ for all $j = 1, \ldots, J$ and $k = 1, \ldots, K$.) Note that (12) and (13) can trivially be extended to functions with domains in more than 2 dimensions, such as three-dimensional images consisting of voxels. In each case we obtain fDO_i and vDO_i values that we can plot in a FOM, with cutoff value (11).

As an illustration we analyze a dataset containing MRI brain images of 416 subjects aged between 18 and 96 (Marcus et al. 2007), which can be freely accessed at *www.oasis-brains.org*. For each subject several images are provided; we will use the masked atlas-registered gain field-corrected images resampled to 1mm isotropic pixels. The masking has set all non-brain pixels to an intensity value of zero. The provided images are already normalized, meaning that the size of the head is exactly the same in each image. The images have 176 by 208 pixels, with grayscale values between 0 and 255. All together we thus have 416 observed images Y_i containing univariate intensity values $Y_i(j, k)$, where $j = 1, \ldots, J = 176$ and $k = 1, \ldots, K = 208$.

There is more information in such an image than just the raw values. We can incorporate shape information by computing the gradient in every pixel of the image. The gradient in pixel (j, k) is defined as the two-dimensional vector $\nabla Y_i(j, k) = (\frac{\partial Y_i(j,k)}{\partial j}, \frac{\partial Y_i(j,k)}{\partial k})$ in which the derivatives have to be approximated numerically. In the pixels at the boundary of the brain we compute forward and backward finite differences, and for the other pixels we employ central differences. In the horizontal direction we thus compute one of three expressions:

$$\frac{\partial Y_i(j,k)}{\partial j} = \begin{cases} (-3\,Y_i(j,k) + 4\,Y_i(j+1,k) & \\ \quad - Y_i(j+2,k))/2 & \text{(forward difference)} \\ (Y_i(j+1,k) - Y_i(j-1,k))/2 & \text{(central difference)} \\ (Y_i(j-2,k) - 4\,Y_i(j-1,k) & \\ \quad + 3\,Y_i(j,k))/2 & \text{(backward difference)} \end{cases}$$

depending on where the pixel is located. The derivatives in the vertical direction are computed analogously.

Incorporating these derivatives yields a dataset of dimensions $416 \times 176 \times 208 \times 3$, so the final $Y_i(j, k)$ are trivariate. For each subject we thus have three data matrices which represent the original MRI image and its derivatives in both directions. Figure 12 shows these three matrices for subject number 387.

The functional DO of an MRI image Y_i is given by (12):

$$\mathrm{fDO}(Y_i; \boldsymbol{Y}) = \frac{1}{176 \times 208} \sum_{j=1}^{176} \sum_{k=1}^{208} \mathrm{DO}(Y_i(j,k); \boldsymbol{Y}(j,k)) W_{jk},$$

where $\mathrm{DO}(Y_i(j, k); \boldsymbol{Y}(j, k))$ is the DO of the trivariate point $Y_i(j, k)$ relative to the trivariate dataset $\{Y_1(j, k), \ldots, Y_{416}(j, k)\}$. In this example, we have set the weight W_{jk} equal to zero at the grid points that are not part of the brain, shown as the black pixels around it. The grid points inside the brain receive full weight.

Figure 13 shows the resulting FOM, which indicates the presence of several outliers. Image 126 has the highest fDO combined with a relatively low vDO. This suggests a shift outlier, that is, a function whose values are all shifted relative to the majority of the data. Images 29 and 92 have a large fDO in combination with a high vDO, indicating that they have strongly outlying subdomains. Images 108, 188, and 234 have an fDO which is on the high end relative to the dataset, which by itself does not make them outlying. Only in combination with their large vDO are they flagged as outliers. These images have strongly outlying subdomains which are smaller that those of functions 29 and 92. The remaining flagged images are fairly close to the cutoff, meaning they are merely borderline cases.

To find out why a particular image is outlying it is instructive to look at a heatmap of its DO values. In Figure 14 we compare the MRI images (on the left) and the DO heatmaps (on the right)

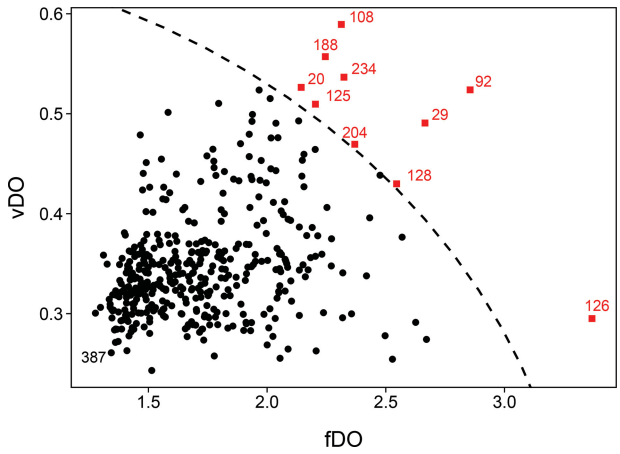

Figure 13. Functional outlier map (FOM) of the MRI dataset, with cutoff curve.

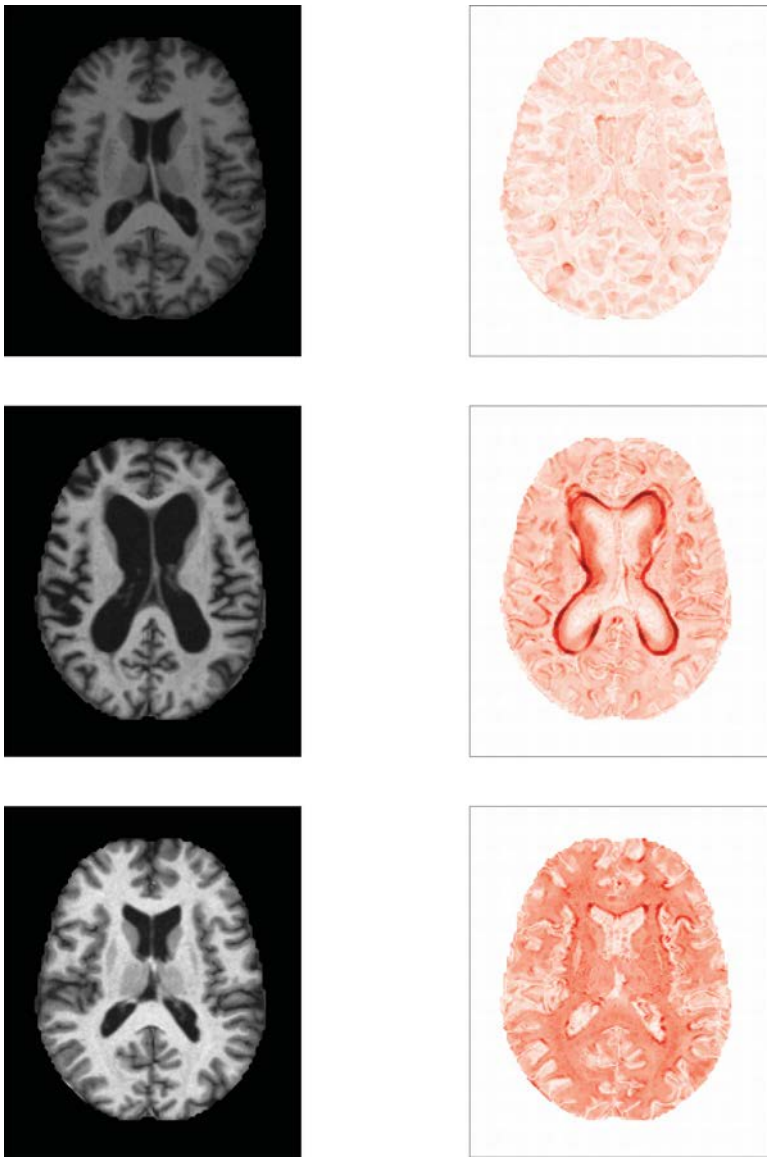

Figure 14. MRI image (left) and DO heatmap (right) of subjects 387 (top), 92 (middle), and 126 (bottom).

of subjects 387, 92, and 126. DO values of 15 or higher received the darkest color. Image 387 has the smallest CFO value, and can be thought of as the most central image in the dataset. As expected, the DO heatmap of image 387 shows very few outlying pixels. For subject 92, the DO heatmap nicely marks the region in which the MRI image deviates most from the majority of the images. Note that the boundaries of this region have the highest outlyingness. This is due to including the derivatives in the analysis, as they emphasize the pixels at which the grayscale intensity changes. The DO heatmap of subject 126 does not show any extremely outlying region but has a rather high outlyingness over the whole domain, which explains its large fDO and regular vDO value. The actual MRI image to its left is globally lighter than the others, confirming that it is a shift outlier.

5. Application to Video

We analyze a surveillance video of a beach, filmed with a static camera (Li et al. 2004). This dataset can be found at *http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html* and consists of 633 frames.

The first 8 seconds of the video show a beach with a tree, as in the leftmost panel of Figure 15. Then a man enters the screen from the left (second panel), disappears behind the tree (third panel), and then reappears to the right of the tree and

Figure 15. Frames number 100, 487, 491, and 500 from the video dataset.

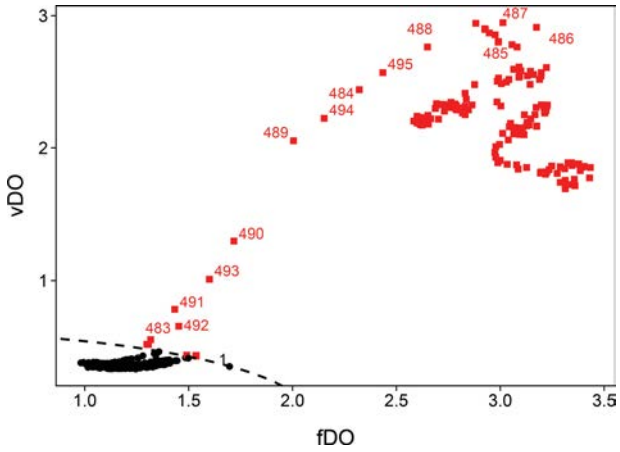

Figure 16. Functional outlier map of the video data.

stays on screen until the end of the video. The frames have 160×128 pixels and are stored using the RGB (Red, Green and Blue) color model, so each frame corresponds to three matrices of size 160×128. Overall we have 633 frames Y_i containing trivariate $Y_i(j,k)$ for $j = 1, \ldots, J = 160$ and $k = 1, \ldots, K = 128$.

Computing the fDO (12) in this dataset is time consuming since we have to execute the projection pursuit algorithm (5) in $\mathbb{R}^3$ for each pixel, so $160 \times 128 = 20{,}480$ times. The entire computation took 25 minutes on a laptop. Therefore we switch to an alternative computation. We define the componentwise DO of a d-variate point $\boldsymbol{y}$ relative to a d-variate sample $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\}$ as

$$\mathrm{CDO}(\boldsymbol{y}; \boldsymbol{Y}) = \sqrt{\sum_{h=1}^{d} \mathrm{DO}(y_h; Y_h)^2}, \tag{14}$$

where $\mathrm{DO}(y_h; Y_h)$ is the univariate DO of the h-th coordinate of $\boldsymbol{y}$ relative to the h th coordinate of $\boldsymbol{Y}$. Analyzing the video data with this componentwise procedure took 15 seconds, so it is a 100 times faster than with projection pursuit, and it produced almost the same FOM. Figure 16 shows the FOM obtained from the CDO computation.

The first 480 frames, which depict the beach and the tree with only the ocean surface moving slightly, are found at the bottom left part of the FOM. They fall inside the dashed curve that separates the regular frames from the outliers. At frame 483 the man enters the picture, making the standard deviation of the DO rise

Figure 17. Left: Frames 100, 487, 491, and 500 from the video. Middle: DO heatmaps of these frames. Right: FOM with blue marker at the position of the frame.

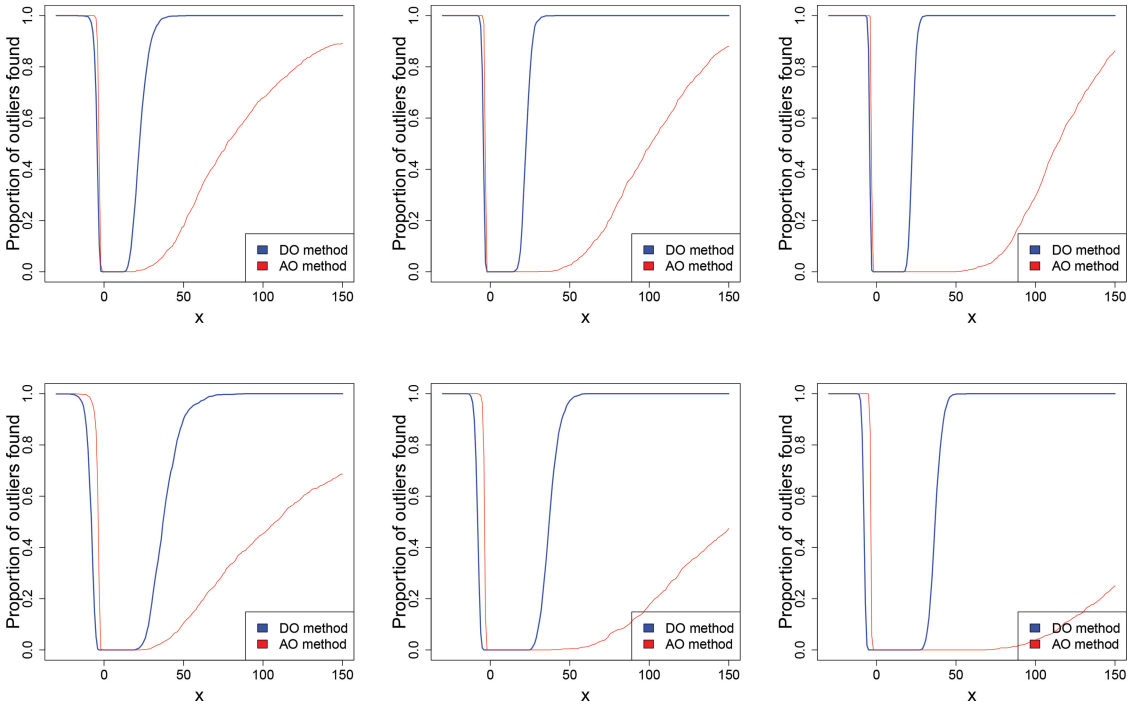

Figure 18. Percentage of outliers found in univariate lognormal samples of size $n = 200$ (left), $n = 500$ (middle), and $n = 1000$ (right), with 10% (top) and 15% (bottom) of outliers in x.

slightly. The fDO increases more slowly, as the fraction of the pixels covered by the man is still low at this stage. This frame can thus be seen as locally outlying. The subsequent frames 484–487 have very high fDO and vDO. In them the man is clearly visible between the left border of the frame and the tree, so these frames have outlying pixels in a substantial part of their domain. Frames 489–492 see the man disappear behind the tree, so the fDO goes down as the fraction of outlying pixels decreases. From frame 493 onward the man reappears to the right of the tree and stays on screen until the end. These frames contain many outlying pixels, yielding points in the upper right part of the FOM.

In the FOM we also labeled frame 1, which lies close to the outlyingness border. Further inspection indicated that this frame is a bit lighter than the others, which might be due to the initialization of the camera at the start of the video.

In addition to the FOM we can draw DO heatmaps of the individual frames. For frames 100, 487, 491, and 500, Figure 17 shows the raw frame on the left, the DO heatmap in the middle and the FOM on the right, in which a blue circle marks the position of the frame. In this figure we can follow the man's path in the FOM, while the DO heatmaps show exactly where the man is in those frames. We have created a video in which the raw frame, the DO heatmap and the FOM evolve alongside each other. It can be downloaded from *http://wis.kuleuven.be/stat/robust/papers-since-2010*.

6. Simulation Study

We would also like to study the performance of the DO when the data generating mechanism is known, and compare it with the AO measure proposed by Brys, Hubert, and Rousseeuw (2005) and studied by Hubert and Van der Veeken (2008) and Hubert, Rousseeuw, and Segaert (2015). For this we carried out an extensive simulation study, covering univariate as well as multivariate and functional data.

In the univariate case we generated $m = 1000$ standard lognormal samples of size $n = \{200, 500, 1000\}$ with 10% and 15% of outliers at the position x, which may be negative. Figure 18 shows the effect of the contamination at x on our figure of merit, the percentage of outliers flagged (averaged over the m replications).

In the direction of the short left tail of the lognormal distribution we see that the adjusted outlyingness AO flags about the same percentage of outliers as the DO. But the AO is much slower in flagging outliers in the direction of the long right tail of the lognormal. This is due to the relatively high explosion bias of the scale used in the denominator of the AO for points to the right of the median. The DO flags outliers to the right of the median much faster, due to its lower explosion bias.

We have also extended the multivariate simulation of AO in Hubert and Van der Veeken (2008). Our simulation consists of $m = 1000$ samples in dimensions $d = \{2, 5, 10\}$ and with sample sizes $n = \{200, 500, 1000\}$. The clean data were generated from the multivariate skew normal distribution (Azzalini and Dalla Valle 1996) with density $f(\boldsymbol{y}) = 2\phi_d(\boldsymbol{y})\Phi(\boldsymbol{\alpha}^T\boldsymbol{y})$ where Φ is the standard normal cdf, ϕ_p is the d-variate standard normal density, and $\boldsymbol{\alpha}$ is a d-variate vector which regulates the shape. In our simulations $\boldsymbol{\alpha}$ is a vector with entries equal to 10 or 4. For $d = 2$ we used $\boldsymbol{\alpha} = (10, 4)^T$, for $d = 5$ we put $\boldsymbol{\alpha} = (10, 10, 4, 4, 4)^T$, and for $d = 10$ we took $\boldsymbol{\alpha} = (10, 10, 10, 10, 10, 4, 4, 4, 4, 4)^T$. To this we added 10% of contamination with a normal distribution $N(\boldsymbol{x}, I_d/20)$ around the point $\boldsymbol{x} = (x, \ldots, x)^T$, where x is on the horizontal axis of Figure 19. In $d = 2$ dimensions we see that AO flags the outliers a bit faster in the direction of the shortest tail, but slower in the direction of the longest tail. The latter is similar to what we saw for univariate data, due to the higher explosion bias of the scale used (implicitly) in the AO. When both the dimension d and the sample size n go up, the DO and AO methods give more similar results. This is because, in most directions, the scales

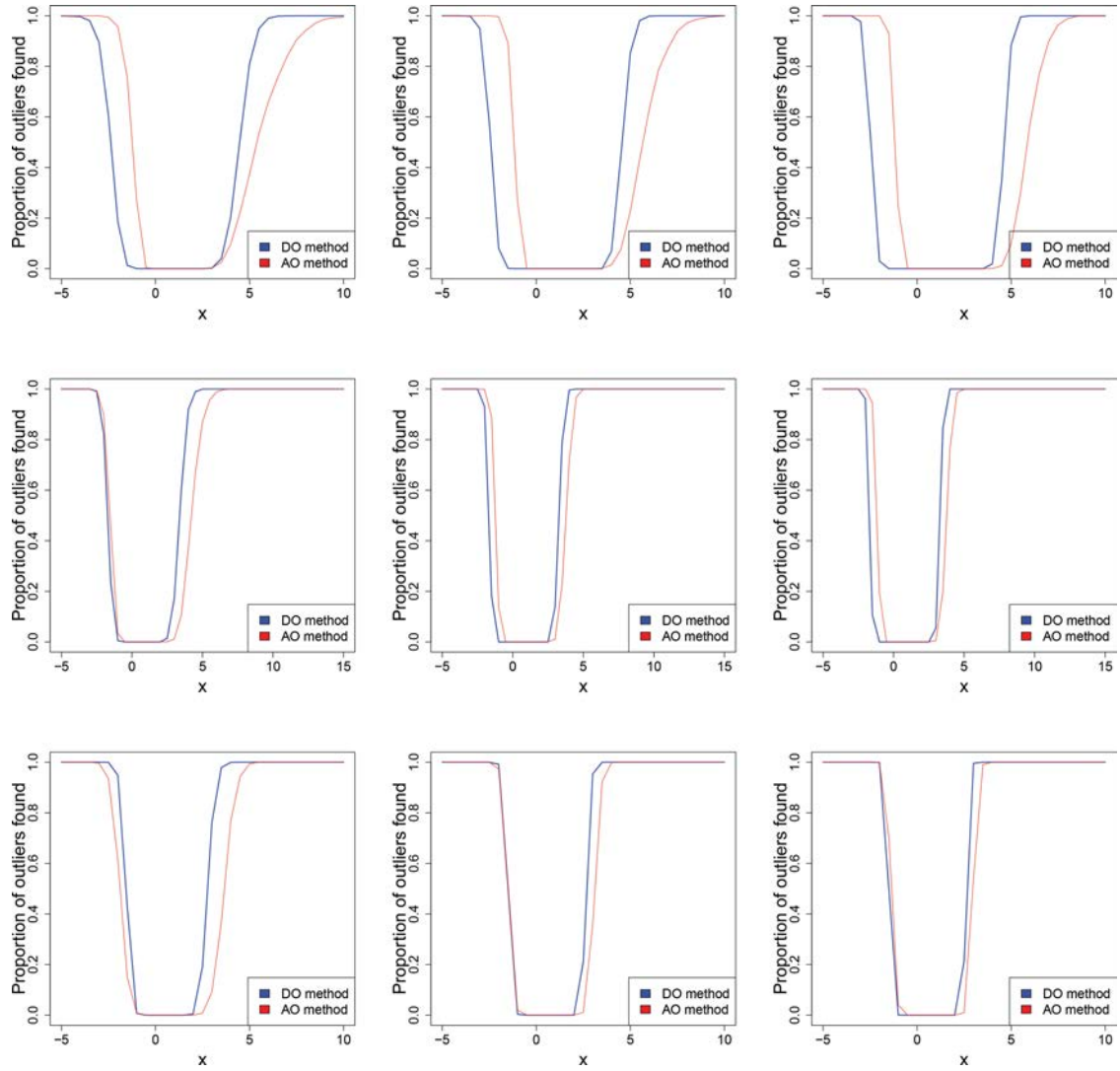

Figure 19. Percentage of outliers found in multivariate skew normal samples of size $n = 200$ (left), $n = 500$ (middle), and $n = 1000$ (right), with 10% of outliers around $\boldsymbol{x} = (x, \ldots, x)^T$, in dimensions $d = 2$ (top), $d = 5$ (middle), and $d = 10$ (bottom).

s_a and s_b of the projected data get closer to each other. This is because the projections of the good data (i.e., without the outliers) tend to become more Gaussian as the dimension d and the sample size n go up, as shown by Diaconis and Freedman (1984) for random directions uniformly distributed on the unit sphere and under moment conditions on the data distribution.

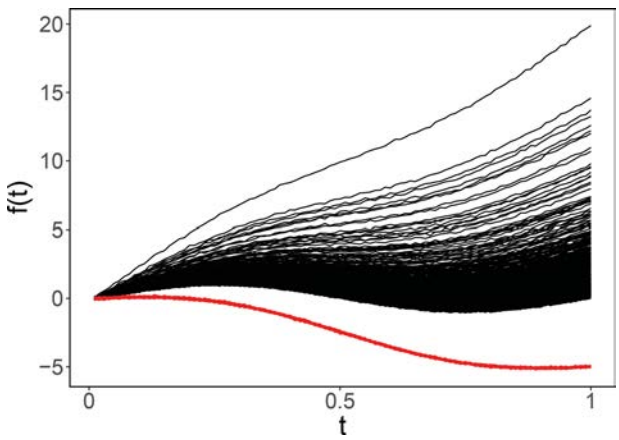

Figure 20. $n = 1000$ generated functions with 10% contamination.

We also carried out a simulation with functional data. We have generated $m = 1000$ samples of $n = \{200, 500, 1000\}$ functions of the form

$$f_i(t) = \sin(2\pi t) + tL_i + \varepsilon_i(t) \quad \text{for} \quad 0 \leqslant t \leqslant 1, \tag{15}$$

where $\ln(L_i) \sim N(0, 1)$ and $\varepsilon_i(t) \sim N(0, (\frac{1}{20})^2)$. That is, the base function is the sine and we add different straight lines of which the slopes are generated by a lognormal distribution. We then replace 10% of the functions by contaminated ones, which are generated from (15) but where L_i is taken higher or lower than what one would expect under the lognormal model. Figure 20 shows such a generated dataset of size $n = 1000$, with outlying functions (with negative L_i) in red.

In the simulation we used a single slope L for the 10% of contaminated curves, and this L is shown on the horizontal axis in Figure 21. When the outlying functions lie below the regular ones (i.e., for negative L), we see that the DO and AO behave similarly. On the other hand, when the outlying functions lie above the regular ones (i.e., in the direction of the long tail), the AO is much slower to flag them than DO.

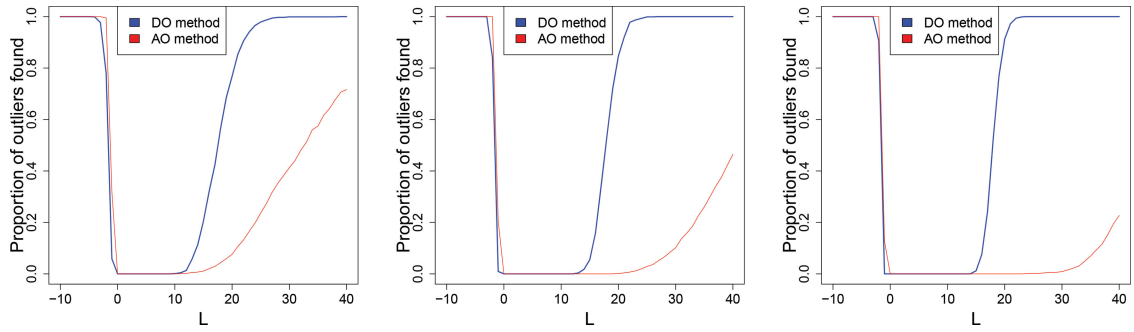

Figure 21. Percentage of outliers found in functional samples of size $n = 200$ (left), $n = 500$ (middle), and $n = 1000$ (right), with 10% of contaminated curves with slope L.

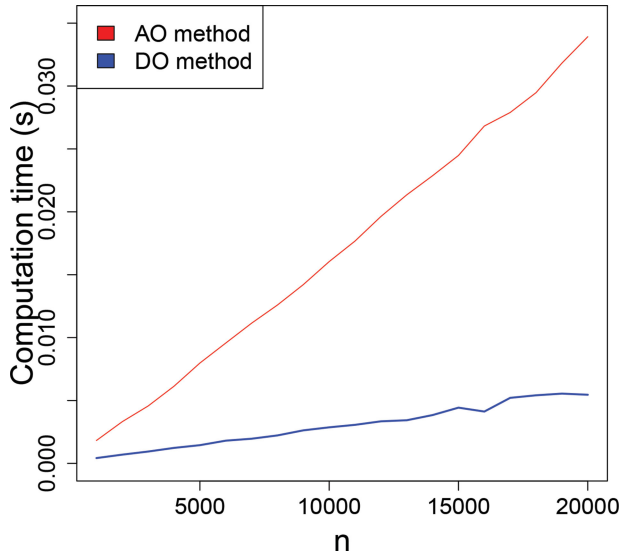

Figure 22. Average computation time of DO and AO as a function of sample size.

These simulations together suggest that the DO outperforms AO in directions where the uncontaminated data has a longer tail, while performing similarly in the other directions.

Note that the DO requires only $\mathcal{O}(n)$ computation time per direction, which is especially beneficial for functional data with a large domain. In particular, DO is much faster than AO which requires $\mathcal{O}(n\log(n))$ operations. Figure 22 shows the average computation time (in seconds) of both measures as a function of the sample size n, for $m = 1000$ samples from the standard normal. The AO time is substantially above the DO time.

7. Conclusion

The notion of directional outlyingness (DO) is well-suited for skewed distributions. It has good robustness properties, and lends itself to the analysis of univariate, multivariate, and functional data, in which both the domain and the function values can be multivariate. Rough cutoffs for outlier detection are available. The DO is also a building block of several graphical tools like DO heatmaps, DO contours, and the functional outlier map (FOM). These proved useful when analyzing spectra, MRI images, and surveillance video. In the MRI images we added gradients to the data to reflect shape/spatial information. In video data we could also add the numerical derivative in the time direction. In our example this would make the frames six-dimensional, but the componentwise DO in (14) would remain fast to compute.

8. Available Software

The methods described in this article are available in R (R core team 2016), in the package *mrfDepth* (*https://CRAN.R-project.org/package=mrfDepth*). An R script and the data sets for reproducing the examples in the article are available from our website *http://wis.kuleuven.be/stat/robust/software*.

Appendix

Proof of Lemma 1(i). Let $\mu \in \mathbb{R}$ be fixed. For the function ρ_c we have that

$$\begin{aligned} & t^2 \int_\mu^\infty \rho_c\left(\frac{x-\mu}{t}\right) dF(x) \\ &= \int_\mu^\infty \left\{ \left(\frac{x-\mu}{c}\right)^2 \mathbb{1}_{|\frac{x-\mu}{t}|\le c} + t^2 \mathbb{1}_{|\frac{x-\mu}{t}|>c} \right\} dF(x) \\ &= \int_0^\infty \left\{ \left(\frac{u}{c}\right)^2 \mathbb{1}_{0\le u\le ct} + t^2 \mathbb{1}_{ct<u} \right\} dF(\mu+u) \end{aligned}$$

For all $u \geqslant 0$ it holds that $(\frac{u}{c})^2 \mathbb{1}_{0\leqslant u\leqslant ct} + t^2 \mathbb{1}_{ct<u}$ is nondecreasing in t, and even strictly increasing in t at large enough u. This proves (i) since $f(x) > 0$ in all x. □

Proof of Lemma 1(ii). Fix $\sigma > 0$. It follows from the Leibniz integral rule that

$$\frac{\partial}{\partial t}\left\{\sigma^2 \int_t^\infty \rho_c\left(\frac{x-t}{\sigma}\right) dF(x)\right\} = -\sigma \int_t^\infty \rho_c'\left(\frac{x-t}{\sigma}\right) dF(x)$$

because $\rho_c(0) = 0$. Note now that

$$\begin{aligned} \rho_c'\left(\frac{x-t}{\sigma}\right) &> 0 \quad \text{for } t \leqslant x < t + \sigma c \\ \rho_c'\left(\frac{x-t}{\sigma}\right) &= 0 \quad \text{for } x > t + \sigma c. \end{aligned}$$

This implies that $\frac{\partial}{\partial t}\{\sigma^2 \int_t^\infty \rho_c(\frac{x-t}{\sigma}) dF(x)\} < 0$ for all t. □

Proof of Proposition 1. Let $0 < \varepsilon < 0.25$ be fixed and let $F_{\varepsilon,H}$ be a minimizing distribution, that is,

$$\inf_{F_{\varepsilon,G}\in\mathcal{F}_{\varepsilon,G}} (\mathrm{s}_a(F_{\varepsilon,G})) = \mathrm{s}_a(F_{\varepsilon,H})$$

with $F_{\varepsilon,G} = (1-\varepsilon)F + \varepsilon G$. Inserting the contaminated distribution $F_{\varepsilon,H}$ into $\mathrm{s}_a(F_{\varepsilon,H})$ yields the scale

$$\begin{aligned} & \frac{\mathrm{s}_{o,a}^2(F_{\varepsilon,H})}{\alpha} \left\{ (1-\varepsilon) \int_{\mathrm{med}(F_{\varepsilon,H})}^\infty \rho_c\left(\frac{x-\mathrm{med}(F_{\varepsilon,H})}{\mathrm{s}_{o,a}(F_{\varepsilon,H})}\right) dF(x) \right. \\ & \left. + \varepsilon \int_{\mathrm{med}(F_{\varepsilon,H})}^\infty \rho_c\left(\frac{x-\mathrm{med}(F_{\varepsilon,H})}{\mathrm{s}_{o,a}(F_{\varepsilon,H})}\right) dH(x) \right\}. \end{aligned} \tag{A.1}$$

For simplicity, put

$$W_1(F_{\varepsilon,H}) = \int_{\mathrm{med}(F_{\varepsilon,H})}^{\infty} \rho_c\left(\frac{x-\mathrm{med}(F_{\varepsilon,H})}{\mathrm{s}_{o,a}(F_{\varepsilon,H})}\right) dF(x)$$
$$W_2(F_{\varepsilon,H}) = \int_{\mathrm{med}(F_{\varepsilon,H})}^{\infty} \rho_c\left(\frac{x-\mathrm{med}(F_{\varepsilon,H})}{\mathrm{s}_{o,a}(F_{\varepsilon,H})}\right) dH(x). \quad (A.2)$$

We then have the contaminated scale

$$\frac{\mathrm{s}^2_{o,a}(F_{\varepsilon,H})}{\alpha}\{(1-\varepsilon)W_1(F_{\varepsilon,H}) + \varepsilon W_2(F_{\varepsilon,H})\}. \quad (A.3)$$

Denote by $Q_{2,\varepsilon} = F^{-1}_{\varepsilon,H}(0.5)$ and $Q_{3,\varepsilon} = F^{-1}_{\varepsilon,H}(0.75)$ the median and the third quartile of the contaminated distribution.

For the distribution H it has to hold that $H(Q_{3,\varepsilon}) = 1$ and $\lim_{x\to Q^-_{2,\varepsilon}} H(x) = 0$. This can be seen as follows. Suppose $H(\infty) - H(Q_{3,\varepsilon}) = p \in (0,1]$. Then consider F_{ε,H^*} where

$$H^*(x) = \begin{cases} H(x) + p\Delta(Q_{2,\varepsilon}) & \text{for } x \in (-\infty, Q_{3,\varepsilon}] \\ 1 & \text{else} \end{cases}$$

and denote by $Q^*_{2,\varepsilon}$ and $Q^*_{3,\varepsilon}$ the median and third quartile of F_{ε,H^*}. Note that $Q_{2,\varepsilon} = Q^*_{2,\varepsilon}$ and $Q_{3,\varepsilon} > Q^*_{3,\varepsilon}$. Therefore, we have $\mathrm{s}_{o,a}(F_{\varepsilon,H}) > \mathrm{s}_{o,a}(F_{\varepsilon,H^*})$. It then follows from Lemma 1(i) that $\mathrm{s}_{o,a}(F_{\varepsilon,H})^2(1-\varepsilon)W_1(F_{\varepsilon,H}) > \mathrm{s}_{o,a}(F_{\varepsilon,H^*})^2(1-\varepsilon)W_1(F_{\varepsilon,H^*})$ and $W_2(F_{\varepsilon,H}) > W_2(F_{\varepsilon,H^*}) = 0$. Therefore, $\mathrm{s}_{o,a}(F_{\varepsilon,H})^2\varepsilon W_2(F_{\varepsilon,H}) > \mathrm{s}_{o,a}(F_{\varepsilon,H^*})^2\varepsilon W_2(F_{\varepsilon,H^*})$. It now follows that $\mathrm{s}_a(F_{\varepsilon,H^*}) < \mathrm{s}_a(F_{\varepsilon,H})$, which is a contradiction since H minimizes s_a. Therefore, $H(\infty) - H(Q_{3,\varepsilon}) = 0$. A similar argument can be made to show that $\lim_{x\to Q^-_{2,\varepsilon}} H(x) = 0$. It follows that $H(Q_{3,\varepsilon}) = 1$ and $H(x) = 0$ for all $x < Q_{2,\varepsilon}$, so all the mass of H is inside $[Q_{2,\varepsilon}\ ,\ Q_{3,\varepsilon}]$.

We can now argue that H must have all its mass in $Q_{2,\varepsilon}$. Note that if $H(Q_{3,\varepsilon}) = 1$ and $\lim_{x\to Q^-_{2,\varepsilon}} H(x) = 0$ we have $Q_{2,\varepsilon} = F^{-1}(\frac{1}{2(1-\varepsilon)})$ and $Q_{3,\varepsilon} \in [F^{-1}(\frac{3-4\varepsilon}{4(1-\varepsilon)}), F^{-1}(\frac{3}{4(1-\varepsilon)})]$, depending on $\lim_{x\to Q^-_{3,\varepsilon}} H(x)$. Given that $Q_{2,\varepsilon}$ is fixed, we can minimize $W_1(F_{\varepsilon,H})$ by minimizing $Q_{3,\varepsilon}$. Now $Q_{3,\varepsilon}$ is minimal for $H = \Delta(F^{-1}(\frac{1}{2(1-\varepsilon)}))$ as this yields $Q_{3,\varepsilon} = F^{-1}(\frac{3-4\varepsilon}{4(1-\varepsilon)})$. Note that this choice of H to minimize $Q_{3,\varepsilon}$ is not unique as any H which makes $\lim_{x\to Q^-_{3,\varepsilon}} H(x) = 1$ does the job. Note finally that $W_2(F_{\varepsilon,H})$ is also minimal for $H = \Delta(F^{-1}(\frac{1}{2(1-\varepsilon)}))$ as $\rho_c(t)$ is nondecreasing in $|t|$, and this choice of H yields $W_2(F_{\varepsilon,H}) = 0$.

We now know that $H = \Delta(F^{-1}(\frac{1}{2(1-\varepsilon)}))$ minimizes $\mathrm{s}_a(F_{\varepsilon,H})$. Furthermore, we have $Q_{2,\varepsilon} = F^{-1}(\frac{1}{2(1-\varepsilon)}) = B^+(\varepsilon, \mathrm{med}, F)$ and $Q_{3,\varepsilon} = F^{-1}(\frac{3-4\varepsilon}{4(1-\varepsilon)})$. Therefore, the implosion bias of s_a is

$$B^-(\varepsilon, \mathrm{s}_a, F)^2 = \frac{B^-(\varepsilon, \mathrm{s}_{o,a}, F)^2}{\alpha} \times \left\{(1-\varepsilon)\int_{B^+(\varepsilon,\mathrm{med},F)}^{\infty} \rho_c\left(\frac{x - B^+(\varepsilon,\mathrm{med},F)}{B^-(\varepsilon,\mathrm{s}_{o,a},F)}\right) dF(x)\right\},$$

where

$$B^+(\varepsilon, \mathrm{med}, F) = F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)$$
$$B^-(\varepsilon, \mathrm{s}_{o,a}, F) = \left(F^{-1}\left(\frac{3-4\varepsilon}{4(1-\varepsilon)}\right) - F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right) / \Phi^{-1}(0.75).$$

□

Proof of Proposition 2. Let $0 < \varepsilon < 0.25$ be fixed and let $F_{\varepsilon,H}$ be a maximizing distribution, that is,

$$\sup_{F_{\varepsilon,G}\in\mathcal{F}_{\varepsilon,G}} (\mathrm{s}_a(F_{\varepsilon,G})) = \mathrm{s}_a(F_{\varepsilon,H})$$

with $F_{\varepsilon,G} = (1-\varepsilon)F + \varepsilon G$. Inserting the contaminated distribution $F_{\varepsilon,H}$ into $\mathrm{s}_a(F_{\varepsilon,H})$ yields the scale (A.1), which can be rewritten as in (A.2) and (A.3).

For the distribution H it has to hold that $H(Q_{3,\varepsilon}) = \lim_{x\to Q^-_{2,\varepsilon}} H(x)$. This can be seen as follows. Suppose $H(Q_{3,\varepsilon}) - \lim_{x\to Q^-_{2,\varepsilon}} H(x) = p \in (0,1]$. Now put $e = B^+(\varepsilon, \mathrm{med}, F) + cB^+(\varepsilon, \mathrm{s}_{o,a}, F)$ and consider the distribution F_{ε,H^*} where

$$H^*(x) = \begin{cases} H(x) & \text{for } x \in (-\infty, Q_{2,\varepsilon}) \\ \lim_{x\to Q^-_{2,\varepsilon}} H(x) & \text{for } x \in [Q_{2,\varepsilon}, Q_{3,\varepsilon}] \\ H(x) - p + p\Delta(e) & \text{for } x \in (Q_{3,\varepsilon}, \infty) \end{cases}$$

and denote by $Q^*_{2,\varepsilon}$ and $Q^*_{3,\varepsilon}$ the median and third quartile of F_{ε,H^*}. Note that $Q_{2,\varepsilon} = Q^*_{2,\varepsilon}$ and $Q_{3,\varepsilon} < Q^*_{3,\varepsilon}$. Therefore $\mathrm{s}_{o,a}(F_{\varepsilon,H}) < \mathrm{s}_{o,a}(F_{\varepsilon,H^*})$ and thus $\mathrm{s}_{o,a}(F_{\varepsilon,H})^2(1-\varepsilon)W_1(F_{\varepsilon,H}) < \mathrm{s}_{o,a}(F_{\varepsilon,H^*})^2(1-\varepsilon)W_1(F_{\varepsilon,H^*})$ because of Lemma 1(i). Furthermore, $W_2(F_{\varepsilon,H}) < W_2(F_{\varepsilon,H^*})$ because $\rho_c(t)$ is nondecreasing in $|t|$, thus $\mathrm{s}_{o,a}(F_{\varepsilon,H})^2\varepsilon W_2(F_{\varepsilon,H}) < \mathrm{s}_{o,a}(F_{\varepsilon,H^*})^2\varepsilon W_2(F_{\varepsilon,H^*})$. It now follows that $\mathrm{s}_a(F_{\varepsilon,H^*}) > \mathrm{s}_a(F_{\varepsilon,H})$, which is a contradiction since H maximizes s_a. Therefore $H(Q_{3,\varepsilon}) = \lim_{x\to Q^-_{2,\varepsilon}} H(x)$, so H has no mass inside $[Q_{2,\varepsilon}\ ,\ Q_{3,\varepsilon}]$.

Without loss of generality we can thus assume that H is of the form $H = d\Delta(e_1) + (1-d)\Delta(e_2)$ with $e_1 = B^-(\varepsilon, \mathrm{med}, F) - cB^+(\varepsilon, \mathrm{s}_{o,a}, F)$ and $e_2 = B^+(\varepsilon, \mathrm{med}, F) + cB^+(\varepsilon, \mathrm{s}_{o,a}, F)$ where $d \in [0,1]$. This choice of e_1 and e_2 is not unique but it maximizes $\mathrm{s}_a(F_{\varepsilon,H})$ because $\rho_c(t)$ is nondecreasing in $|t|$. Inserting the distribution $F_d := F_{\varepsilon,H}$ yields

$$\mathrm{s}^2_a(F_d) = \frac{\mathrm{s}^2_{o,a}(F_d)}{\alpha}\left\{(1-\varepsilon)\int_{Q_{2,d}}^{\infty} \rho_c\left(\frac{x-Q_{2,d}}{\mathrm{s}_{o,a}(F_d)}\right) dF(x) + \varepsilon(1-d)\right\}, \quad (A.4)$$

where $Q_{2,d} = F^{-1}(\frac{1-2d\varepsilon}{2(1-\varepsilon)})$, $Q_{3,d} = F^{-1}(\frac{3-4d\varepsilon}{4(1-\varepsilon)})$, and $\mathrm{s}_{o,a}(F_d) = (Q_{3,d} - Q_{2,d})/\Phi^{-1}(0.75)$. Note that this expression depends on d but no longer on e_1 and e_2. We will show that this expression is maximized for $d = 0$.

First we show that $\mathrm{s}_{o,a}(F_d)$ is maximized for $d = 0$. Let

$$g(d) := \mathrm{s}_{o,a}(F_d) = (Q_{3,d} - Q_{2,d})/\Phi^{-1}(0.75) \quad (A.5)$$

for any $d \in [0,1]$. Note that $\xi = \frac{3-4\varepsilon d}{4(1-\varepsilon)} - \frac{2-4\varepsilon d}{4(1-\varepsilon)} = \frac{1}{4(1-\varepsilon)}$ does not depend on d. Therefore, we can write $g(d) = (F^{-1}(\upsilon+\xi) - F^{-1}(\upsilon))/\Phi^{-1}(0.75)$ where $\upsilon = \frac{2-4\varepsilon d}{4(1-\varepsilon)}$ is a strictly decreasing function of d. Note that we can write $g(d) = (\Phi^{-1}(0.75))^{-1}\int_{\upsilon}^{\upsilon+\xi} \frac{du}{f(F^{-1}(u))}$. The density f is symmetric about some m, and by affine equivariance we can assume $m = 0$ w.l.o.g. Since f is unimodal with $f(x) > 0$ for all x, the function $u \to \frac{1}{f(F^{-1}(u))}$ is strictly decreasing up to its minimum (corresponding to the mode of f) and then strictly increasing. Therefore, $g(d)$ is maximal for υ as large as possible, that is, for $d = 0$. In that case, we have $\upsilon = Q_{2,o} = \frac{2}{4(1-\varepsilon)} > 0.5$.

Next, we maximize

$$h(\sigma, d) := \frac{\sigma^2}{\alpha}\left\{(1-\varepsilon)\int_{Q_{2,d}}^{\infty} \rho_c\left(\frac{x-Q_{2,d}}{\sigma}\right) dF(x) + \varepsilon(1-d)\right\} \quad (A.6)$$

for any fixed $\sigma > 0$. This is equivalent to maximizing

$$\int_q^{\infty} \rho_c\left(\frac{x-q}{\sigma}\right) dF(x) + \frac{\varepsilon}{1-\varepsilon}(1-d), \quad (A.7)$$

where q is such that $F(q) \in [\frac{1-2\varepsilon}{2(1-\varepsilon)}, \frac{1}{2(1-\varepsilon)}] = \frac{1}{2} \pm \frac{\varepsilon}{1-\varepsilon}$. Note that $\frac{\varepsilon}{1-\varepsilon}(1-d) = F(q) + \frac{1-2\varepsilon}{2(1-\varepsilon)}$, where the second term doesn't depend on q. Maximizing (A.7) with respect to q is therefore equivalent to maximizing $\int_q^{q+c\sigma} (\frac{x-q}{c\sigma})^2 dF(x) - \int_q^{q+c\sigma} dF(x)$. Note that this is equal to $\int_0^{c\sigma} (\frac{x}{c\sigma})^2 dF(q+x) - \int_0^{c\sigma} dF(q+x) = \int_0^{c\sigma} \frac{x^2-\sigma^2c^2}{\sigma^2c^2} f(q+x)d(x)$. For all x in $[0, c\sigma]$ it holds that $\frac{x^2-\sigma^2c^2}{\sigma^2c^2} \le 0$, hence the latter integral is maximized by minimizing $f(q+x)$ for all $x \in [0, c\sigma]$. For this q must take on its highest possible value $q = F^{-1}(\frac{1}{2} + \frac{\varepsilon}{1-\varepsilon})$, because we then have $f(q+x) \le f(q_2+x)$ for all x in $[0, c\sigma]$ and all q_2 in $[F^{-1}(\frac{1}{2} - \frac{\varepsilon}{1-\varepsilon}), F^{-1}(\frac{1}{2} + \frac{\varepsilon}{1-\varepsilon})]$. Therefore, (A.6) is maximized for $d = 0$.

We now know that (A.5) and (A.6) satisfy $\max_d g(d) = g(0)$ and $\max_d h(\sigma, d) = h(\sigma, 0)$ for all $\sigma > 0$. By Lemma 1(i), $h(\sigma, 0)$ is increasing in σ. Combining these results yields

$$\begin{aligned}\max_d \mathrm{s}_a^2(F_d) &= \max_d h(g(d), d)\\ &\leqslant \max_{d_1}\max_{d_2} h(g(d_1), d_2)\\ &= \max_{d_1} h(g(d_1), 0)\\ &= h(g(0), 0),\end{aligned}$$

so $\mathrm{s}_a^2(F_d)$ is maximized for $d = 0$. Therefore, $\mathrm{s}_a^2(F_{\varepsilon,H})$ is maximized for $H = \Delta(e_2)$, hence $Q_{2,\varepsilon} = F^{-1}(\frac{1}{2(1-\varepsilon)}) = B^+(\varepsilon, \mathrm{med}, F)$ and $Q_{3,\varepsilon} = F^{-1}(\frac{3}{4(1-\varepsilon)})$. The explosion bias is thus

$$B^+(\varepsilon, \mathrm{s}_a, F)^2 = \frac{\mathrm{s}_{o,a}^2}{\alpha}\left\{(1-\varepsilon)\int_{B^+(\varepsilon,\mathrm{med},F)}^{\infty}\rho_c\left(\frac{x - B^+(\varepsilon, \mathrm{med}, F)}{\mathrm{s}_{o,a}}\right)dF(x) + \varepsilon\right\},$$

where

$$\mathrm{s}_{o,a} = \left\{F^{-1}\left(\frac{3}{4(1-\varepsilon)}\right) - F^{-1}\left(\frac{1}{2(1-\varepsilon)}\right)\right\}\Big/\,\Phi^{-1}(0.75).$$

□

Proof of Proposition 3. Plugging the contaminated distribution $F_{\varepsilon,z} = (1-\varepsilon)F + \varepsilon\Delta z$ into the functional form (4) of s_a yields

$$\mathrm{s}_a^2(F_{\varepsilon,z}) = \frac{\mathrm{s}_{o,a}^2(F_{\varepsilon,z})}{\alpha}\int_{\mathrm{med}(F_{\varepsilon,z})}^{\infty}\rho_c\left(\frac{x - \mathrm{med}(F_{\varepsilon,z})}{\mathrm{s}_{o,a}(F_{\varepsilon,z})}\right)dF_{\varepsilon,z}(x).$$

We take the derivative with respect to ε and evaluate it in $\varepsilon = 0$. Note that $\rho_c(t)$ is not differentiable at $t = c$ and $t = -c$, but as these two points form a set of measure zero this does not affect the integral containing $\rho_c'(t)$. We also use that $\rho_c(0) = 0$ and $\mathrm{IF}(z, \mathrm{s}_a^2, F) = 2\mathrm{s}_a(F)\mathrm{IF}(z, \mathrm{s}_a, F)$ yielding the desired expression. For $F = \Phi$ we have $\mathrm{IF}(z, \mathrm{s}_{o,a}, \Phi) = (\mathbb{1}_{\{[0,\infty)\}}(z)\mathrm{sign}(z - \Phi^{-1}(\frac{3}{4})) + \mathrm{IF}(z, \mathrm{med}, \Phi)\{\phi(0) - 2\phi(\Phi^{-1}(\frac{3}{4}))\})/(2\phi(\Phi^{-1}(\frac{3}{4})))$. □

Proof of Proposition 4. We show the proof for $x > \mathrm{med}(F)$, the other case being analogous. Plugging the contaminated distribution $F_{\varepsilon,z} = (1-\varepsilon)F + \varepsilon\Delta z$ into $\mathrm{DO}(x, F) = (x - \mathrm{med}(F))/\mathrm{s}_a(F)$ yields

$$\begin{aligned}\mathrm{IF}(z, \mathrm{DO}(x), F) &= \frac{\partial}{\partial\varepsilon}(\mathrm{DO}(x, F_{\varepsilon,z}))\Big|_{\varepsilon=0} = \frac{1}{\mathrm{s}_a^2(F_{\varepsilon,z})}\Big(-\frac{\partial}{\partial\varepsilon}\Big(\mathrm{med}(F_{\varepsilon,z})\Big)\\ &\quad\times \mathrm{s}_a(F_{\varepsilon,z}) - \frac{\partial}{\partial\varepsilon}(\mathrm{s}_a(F_{\varepsilon,z}))(x - \mathrm{med}(F_{\varepsilon,z}))\Big)\Big|_{\varepsilon=0}\\ &= \frac{1}{\mathrm{s}_a^2(F)}(-\mathrm{IF}(z, \mathrm{med}, F)\mathrm{s}_a(F)\\ &\quad - \mathrm{IF}(z, \mathrm{s}_a, F)(x - \mathrm{med}(F))).\end{aligned}$$

□

Acknowledgments

This research has been supported by projects of Internal Funds KU Leuven. The authors are grateful for interesting discussions with Pieter Segaert.

Funding

Onderzoeksraad, KU Leuven.

References

Arribas-Gil, A., and Romo, J. (2014), "Shape Outlier Detection and Visualization for Functional Data: The Outliergram," *Biostatistics*, 15, 603–619. [345]

Azzalini, A., and Dalla Valle, A. (1996), "The Multivariate Skew-Normal Distribution," *Biometrika*, 83, 715–726. [355]

Berrendero, J., Justel, A., and Svarc, M. (2011), "'Principal Components for Multivariate Functional Data," *Computational Statistics & Data Analysis*, 55, 2619–2634. [345]

Brys, G., Hubert, M., and Rousseeuw, P. J. (2005), "A Robustification of Independent Component Analysis," *Journal of Chemometrics*, 19, 364–375. [345,346,355]

Brys, G., Hubert, M., and Struyf, A. (2004), "A Robust Measure of Skewness," *Journal of Computational and Graphical Statistics*, 13, 996–1017. [346]

Diaconis, P., and Freedman, D. (1984), "Asymptotics of Graphical Projection Pursuit," *The Annals of Statistics*, 12, 793–815. [356]

Donoho, D. (1982), *Breakdown Properties of Multivariate Location Estimators*, Ph.D. dissertation, Department of Statistics, Harvard University. [345]

Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2008), "Outlier Detection in Functional Data by Depth Measures, With Application to Identify Abnormal NO_x Levels," *Environmetrics*, 19, 331–345. [345]

Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, New York, NY: Springer. [345]

Hampel, F., Ronchetti, E., Rousseeuw, P., and Stahel, W. (1986), *Robust Statistics: The Approach Based on Influence Functions*, New York: Wiley. [347]

Hand, D., Lunn, A., McConway, A., and Ostrowski, E. (1994), *A Handbook of Small Data Sets*, London, UK: Chapman and Hall. [348]

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015), "Multivariate Functional Outlier Detection" (with discussion), *Statistical Methods & Applications*, 24, 177–246. [345,349,350,355]

Hubert, M., and Van der Veeken, S. (2008), "Outlier Detection for Skewed Data," *Journal of Chemometrics*, 22, 235–246. [348,355]

Hyndman, R., and Shang, H. (2010), "Rainbow Plots, Bagplots, and Boxplots for Functional Data," *Journal of Computational and Graphical Statistics*, 19, 29–45. [345]

Lemberge, P., De Raedt, I., Janssens, K., Wei, F., and Van Espen, P. (2000), "Quantitative Z-Analysis of 16th–17th Century Archaeological Glass Vessels Using PLS Regression of EPXMA and μ-XRF Data," *Journal of Chemometrics*, 14, 751–763. [348]

Li, L., Huang, W., Gu, I. Y.-H., and Tian, Q. (2004), "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," *IEEE Transactions on Image Processing*, 13, 1459–1472. [353]

Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., and Buckner, R. (2007), "Open Access Series of Imaging Studies (OASIS): Cross-Sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults," *Journal of Cognitive Neuroscience*, 19, 1498–1507. [352]

Martin, R., and Zamar, R. (1993), "Bias Robust Estimation of Scale," *The Annals of Statistics*, 21, 991–1017. [346]

Pigoli, D., and Sangalli, L. (2012), "Wavelets in Functional Data Analysis: Estimation of Multidimensional Curves and Their Derivatives," *Computational Statistics & Data Analysis*, 56, 1482–1498. [345]

R Core Team (2016), "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, Vienna, Austria. Available at: *https://www.R-project.org/* [xxxx]

Ramsay, J., and Silverman, B. (2005), *Functional Data Analysis* (2nd ed.), New York: Springer. [345]

Rousseeuw, P., and Croux, C. (1994), "The Bias of k-Step M-Estimators," *Statistics & Probability Letters*, 20, 411–420. [348]

Stahel, W. (1981), *Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*, Ph.D. dissertation. Switzerland: ETHZürich. [345]

Sun, Y., and Genton, M. (2011), "Functional Boxplots," *Journal of Computational and Graphical Statistics*, 20, 316–334. [345]