

*Submitted to Bernoulli*

# M-estimators of location for functional data

BEATRIZ SINOVA<sup>1,3,\*</sup> GIL GONZÁLEZ-RODRÍGUEZ<sup>1</sup> and STEFAN VAN AELST<sup>2,3</sup>

<sup>1</sup>*Departamento de Estadística, I.O. y D.M., Facultad de Ciencias  
Universidad de Oviedo, 33071 Oviedo, Spain. E-mail: [sinovabeatriz@uniovi.es](mailto:sinovabeatriz@uniovi.es)  
(Corresponding author)*

<sup>2</sup>*Department of Mathematics, KU Leuven, 3001 Leuven, Belgium.*

<sup>3</sup>*Department of Applied Mathematics, Computer Science and Statistics,  
Ghent University, 9000 Gent, Belgium.*

**Abstract** M-estimators of location are widely used robust estimators of the center of univariate or multivariate real-valued data. This paper aims to study M-estimates of location in the framework of functional data analysis. To this end, recent developments for robust nonparametric density estimation by means of M-estimators are considered. These results can also be applied in the context of functional data analysis and allow to state conditions for the existence and uniqueness of location M-estimates in this setting. Properties of these functional M-estimators are investigated. In particular, their consistency is shown and robustness is studied by means of their breakdown point and their influence function. The finite-sample performance of the M-estimators is explored by simulation. The M-estimators are also empirically compared to trimmed means for functional data.

*Keywords:* functional data, functional data metric, Hampel loss, Huber loss, Tukey loss, M-estimates of location, statistical robustness, trimmed means.

## 1. Introduction

The importance of functional data is vastly increasing. Functional data are involved in many real-life problems, such as e.g. the search for patterns of human brain activity (see e.g. Rosen, B.R. *et al.*, 1998), the modeling of human movements applied to humanoid robotics (see e.g. Sreenivasa, M. *et al.*, 2012), the description of food properties (see e.g. Jacques, J. and Preda, C., 2014) or the cerebral white matter tracts of subjects with a certain illness (see e.g. Ciarleglio, A. and Ogden, R.T., 2016). As a result, functional data analysis also receives a lot of attention lately — see e.g. Ramsay, J. and Silverman, B.W. (2005) for a general perspective on the topic and Ferraty, F. and Vieu, P. (2006) for a non-parametric approach.

Similarly as for finite dimensional data, it has been shown that the functional mean is highly sensitive to the presence of outliers and other anomalies. Therefore, several robust estimators of functional location have already been proposed in the literature, such as functional trimmed means (Cuesta-Albertos, J.A. and Fraiman, R., 2007) and functional

location estimators based on depth functions in functional Hilbert spaces (see Cuesta-Albertos, J.A. and Nieto-Reyes, A., 2008; Cuevas, A. and Fraiman, R., 2009; López-Pintado, S. and Romo, J., 2009, 2011, among others), but the interesting alternative of M-estimation has not been considered yet, as far as the authors know (see, for example Fraiman, R. and Svarc, M., 2013).

One of the first approaches to estimate location robustly consists of M-estimators as introduced by Huber, P.J. (1964). M-estimators aim to reduce the large influence of outliers on the standard least squares/maximum likelihood estimators. For this purpose, the key idea is to replace the square loss function by a less rapidly increasing loss function. For univariate location estimation, M-estimators can be seen as intermediaries between the sample mean and the sample median, which can combine high robustness with high efficiency. M-estimators are well-established robust methods in multivariate data analysis (Huber, P.J., 1981; Hampel, F.R. *et al.*, 1986; Maronna, R.A. *et al.*, 2006). Recently, the extension of robust methods to more complex data settings has received increasing attention (see e.g. Filzmoser, P. and Todorov, V., 2013). Some recent developments include Christmann, A. and Steinwart, I. (2007); Steinwart, I. and Christmann, A. (2008); Domingues, M.A.O. *et al.* (2010); D’Urso, P. *et al.* (2011, 2015); Hu, Q. *et al.* (2011); Winkler, R. *et al.* (2011); Sinova, B. *et al.* (2012, 2015); Fritz, H. *et al.* (2013); Szilágyi, L. (2013); Chichignoud, M. and Lederer, J. (2014); Claeskens, G. *et al.* (2014); D’Urso, P. and De Giovanni, L. (2014); Hubert, M. *et al.* (2015) and Minsker, S. (2015).

In this paper, the notion of M-estimators is extended to the functional data setting. Recent ideas of Kim, J.S. and Scott, C.D. (2012) in the framework of robust kernel density estimation introduce M-estimators for Hilbert space valued data. These results allow us to study M-estimators of functional location. In Section 2, we define M-estimators of functional location and show that appropriate conditions on the loss function guarantee their existence and uniqueness. Moreover, the M-estimators can be expressed as a weighted average of the functional observations, which allows to compute them via a reweighted least squares algorithm. In Section 3, we show that the functional M-estimators are translation equivariant, but not scale equivariant in general. Moreover, it is shown that M-estimators are strongly consistent and Borel-measurable. Their robustness is investigated by their finite sample breakdown point as well as their influence function. The finite-sample behavior of the estimators is analyzed by a simulation study in Section 4, where we also make a comparison with functional trimmed means. Section 5 illustrates the use of functional location M-estimators by means of a real-life example, while some concluding remarks are given in Section 6.

## 2. Location M-estimators for functional data

Kim, J.S. and Scott, C.D. (2012) have proposed robust nonparametric density estimators by combining kernel density estimation with ideas from standard M-estimation (see also Kim, J.S., 2011; Kim, J.S. and Scott, C.D., 2011). They have interpreted the traditional kernel density estimator based on a radial, positive semi-definite kernel as the sample mean in the associated reproducing kernel Hilbert space, i.e. a Hilbert space of functions

in which pointwise evaluation is a continuous linear functional (see Steinwart, I. and Christmann, A., 2008; González, J. *et al.*, 2014; Ma, J. *et al.*, 2014; Zhang, Z. and Zhang, Y., 2014, for recent work on different problems involving these spaces). To lower the sensitivity of this mean to outliers, Kim, J.S. and Scott, C.D. (2012) suggest to estimate the center robustly via M-estimators, yielding a robust kernel density estimator. Although Kim and Scott have developed their ideas for reproducing kernel Hilbert spaces, they have also generalized the results to other Hilbert spaces within the setting of kernel density estimation. Moreover, their results remain valid when they are applied to general random variables taking values on a Hilbert space of functions. We exploit this generality to adapt their results to M-estimation of functional location.

As Hsing, T. and Eubank, R. (2015) point out, there are two different perspectives when considering functional data. Functional data could be understood either as sample paths of a stochastic process with smooth mean and covariance functions, or as realizations of Hilbert space valued random variables, which is our approach in this work. Hence, let  $(\Omega, \mathcal{A}, P)$  be a probability space and let  $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$  be a Hilbert space of functions on some set, where  $\|\cdot\|_{\mathbb{H}}$  is the norm associated with the inner product  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ . Let  $X : \Omega \rightarrow \mathbb{H}$  be a Hilbert space valued random variable associated with the probability space (i.e., a measurable function w.r.t.  $\mathcal{A}$  and  $\mathcal{B}_{\|\cdot\|_{\mathbb{H}}}$ , the Borel  $\sigma$ -field generated by the topology induced by  $\|\cdot\|_{\mathbb{H}}$ ).

The functional M-estimators of location are based on a real-valued *loss function*  $\rho$ . We assume that the loss function  $\rho$  satisfies the following conditions.

C.1  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}$  is continuous and non-decreasing, and  $\rho(0) = 0$ .

These assumptions on  $\rho$  are commonly made, see e.g. Maronna, R.A. *et al.* (2006). Following the ideas of Huber, P.J. (1964, 1981), the M-location value and the associated M-estimator of location for Hilbert space valued random variables are defined as follows.

**Definition 2.1** Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  be a Hilbert space with associated norm  $\|\cdot\|_{\mathbb{H}}$ ,  $X : \Omega \rightarrow \mathbb{H}$  be an associated Hilbert space valued random variable such that  $E(\|X\|_{\mathbb{H}}) < \infty$  and  $\rho$  a loss function satisfying C.1. Then, the corresponding M-location value is (are) the element(s)  $g^M(X)$  in  $\mathbb{H}$  given by

$$g^M(X) = \arg \min_{h \in \mathbb{H}} J(h) = \arg \min_{h \in \mathbb{H}} E[\rho(\|X - h\|_{\mathbb{H}})], \quad (1)$$

if it exists.

**Definition 2.2** Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  be a Hilbert space with associated norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable. Moreover, let  $(X_1, \dots, X_n)$  be a simple random sample from  $X$  and  $\rho$  a loss function satisfying C.1. Then, the M-estimator of location is any Hilbert space valued statistic  $\widehat{g}^M(X_1, \dots, X_n)$  which solves

$$\widehat{g}^M(X_1, \dots, X_n) = \arg \min_{h \in \mathbb{H}} J_n(h) = \arg \min_{h \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^n \rho(\|X_i - h\|_{\mathbb{H}}), \quad (2)$$

if it exists.

Note that some additional assumptions on the probability space and Hilbert space may be required to guarantee the Borel measurability of M-estimators of location (see Proposition 3.3).

An important contribution of Kim, J.S. and Scott, C.D. (2012) is the establishment of conditions to ensure the existence of Hilbert space valued M-estimators, as well as their representation as weighted means (i.e., convex linear combinations) of the sample elements. Notice that those weights also depend on the M-estimators, so this representation directly leads to an iterative reweighted least squares algorithm for the computation of M-estimates. We summarize these results for the special case of functional location M-estimators.

**Theorem 2.1 (Representer theorem)** [adapted from Kim & Scott, 2012]. *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a Hilbert space with associated norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable such that  $E(\|X\|_{\mathbb{H}}) < \infty$ . Moreover, let  $\rho$  be a loss function which satisfies conditions C.1 as well as*

$$\text{C.2 } \rho(x)/x \xrightarrow{x \rightarrow 0} 0,$$

$$\text{C.3 } \rho \text{ is differentiable and } \rho' \text{ and } \phi(x) = \rho'(x)/x \text{ are both continuous and bounded, where we assume that } \phi(0) := \lim_{x \rightarrow 0} \rho'(x)/x \text{ exists and is finite.}$$

Then, if  $E(\|X\|_{\mathbb{H}}) < \infty$  and the M-location value exists, any  $g^M(X)$  can be expressed as

$$g^M(X) = \int_{\Omega} u(X) X dP$$

with  $u(X) = \phi(\|X - g^M(X)\|_{\mathbb{H}}) / \int_{\Omega} \phi(\|X - g^M(X)\|_{\mathbb{H}}) dP$ ,

If  $(X_1, \dots, X_n)$  is a simple random sample from  $X$  and the M-estimator of location exists, any  $\widehat{g^M}(X_1, \dots, X_n)$  can be represented as

$$\widehat{g^M}(X_1, \dots, X_n) = \sum_{i=1}^n u_i X_i,$$

with  $u_i = \phi(\|X_i - \widehat{g^M}(X_1, \dots, X_n)\|_{\mathbb{H}}) / \sum_{j=1}^n \phi(\|X_j - \widehat{g^M}(X_1, \dots, X_n)\|_{\mathbb{H}})$ .

Furthermore, if in addition to C.1 – C.3 it holds that

$$\text{C.4 the function } J \text{ (or } J_n, \text{ respectively) is strictly convex,}$$

then the conditions

- $g^M(X) = \int_{\Omega} u(X) X dP$  (or  $\widehat{g^M}(X_1, \dots, X_n) = \sum_{i=1}^n u_i X_i$ ),
- $u \propto \phi(\|X - \widehat{g^M}(X_1, \dots, X_n)\|_{\mathbb{H}})$  (or  $u_i \propto \phi(\|X_i - \widehat{g^M}(X_1, \dots, X_n)\|_{\mathbb{H}})$ ),
- $E[u] = 1$  (or  $\sum_{i=1}^n u_i = 1$ ),

are sufficient to guarantee that  $g^M(X)$  is the functional M-location value in (1) (or  $\widehat{g^M}(X_1, \dots, X_n)$  is the location M-estimator in (2), respectively).

The standard least squares loss function  $\rho(x) = x^2$  does not satisfy condition C.3, but it is well-known that this loss function does not provide robust solutions. Three common families of loss functions that fulfill conditions C.1-C.3 are Huber (Huber, P.J., 1964, 1981), Tukey (Beaton, A.E. and Tukey, J.W., 1974) and Hampel (Hampel, F.R., 1974) families. Huber's family of loss functions is given by

$$\rho_a^H(x) = \begin{cases} x^2/2 & \text{if } 0 \leq x \leq a \\ a(x - a/2) & \text{if } a < x, \end{cases} \quad (3)$$

with tuning parameter  $a > 0$ . This is a convex, but not strictly convex loss function. The Huber family of loss functions provides a hybrid approach between squared ( $x \leq a$ ) and absolute losses ( $x \geq a$ ). It thus puts less emphasis on large errors compared to the squared error loss function.

Tukey's biweight or bisquare family of loss functions is given by

$$\rho_c^T(x) = \begin{cases} c^2/6 \cdot (1 - (1 - (x/c)^2)^3) & \text{if } 0 \leq x \leq c \\ c^2/6 & \text{if } c < x, \end{cases} \quad (4)$$

with tuning parameter  $c > 0$ . This loss function is not convex anymore and once an observation lies far from the center, with error larger than  $c$  ( $x \geq c$ ), it does not matter anymore how far it actually is, the contribution to the loss does not change anymore. This loss function can thus better cope with extreme outliers.

Hampel's family of loss functions is given by

$$\rho_{a,b,c}(x) = \begin{cases} x^2/2 & \text{if } 0 \leq x < a \\ a(x - a/2) & \text{if } a \leq x < b \\ \frac{a(x-c)^2}{2(b-c)} + a(b+c-a)/2 & \text{if } b \leq x < c \\ a(b+c-a)/2 & \text{if } c \leq x, \end{cases} \quad (5)$$

with nonnegative tuning parameters  $0 < a < b < c$ . This loss function shares with the Tukey function the non convexity and the fact that observations far from the center ( $x \geq c$ ) always contribute in the same way to the loss.

The following proposition generalizes the result in Kim, J.S. and Scott, C.D. (2012) that provides sufficient conditions for the strict convexity of the functions  $J$  and  $J_n$  needed in the second part of Theorem 2.1.

**Proposition 2.2** *The strict convexity of  $J$  (respectively,  $J_n$ ) holds if either*

C.4i  *$\rho$  is strictly convex,*

*or*

C.4ii  *$\rho$  is convex, strictly increasing and  $P$  is not concentrated on a line in  $\mathbb{H}$  (or not all of the observations in the sample  $(X_1, \dots, X_n)$  are collinear in  $\mathbb{H}$ , respectively).*

*Proof.* For condition C.4i, see [Kim, J.S. and Scott, C.D. \(2012\)](#). For C.4ii in case of the function  $J$ , take  $\gamma \in (0, 1)$  and let  $g \neq h$  be two arbitrary elements of  $\mathbb{H}$ . Let  $l$  denote the line passing through  $g$  and  $h$ . Let us consider  $\omega \in \Omega$  such that  $X(\omega)$  does not belong to the line  $l$ . Then,  $X(\omega) - h$  and  $X(\omega) - g$  are not linearly dependent. Therefore, the following triangle inequality is strict:

$$\|\gamma X(\omega) - \gamma g + (1 - \gamma)X(\omega) - (1 - \gamma)h\|_{\mathbb{H}} < \gamma\|X(\omega) - g\|_{\mathbb{H}} + (1 - \gamma)\|X(\omega) - h\|_{\mathbb{H}}.$$

As  $\rho$  is strictly increasing,

$$\rho(\|\gamma X(\omega) - \gamma g + (1 - \gamma)X(\omega) - (1 - \gamma)h\|_{\mathbb{H}}) < \rho(\gamma\|X(\omega) - g\|_{\mathbb{H}} + (1 - \gamma)\|X(\omega) - h\|_{\mathbb{H}}).$$

Since the strict inequality expressed above holds for the set  $\{\omega \in \Omega : X(\omega) \notin l\}$ , which has positive probability by assumption, it holds that

$$J(\gamma g + (1 - \gamma)h) = E[\rho(\|X - (\gamma g + (1 - \gamma)h)\|_{\mathbb{H}})]$$

$$= E[\rho(\|\gamma X - \gamma g + (1 - \gamma)X - (1 - \gamma)h\|_{\mathbb{H}})] < E[\rho(\gamma\|X - g\|_{\mathbb{H}} + (1 - \gamma)\|X - h\|_{\mathbb{H}})].$$

Using the convexity of  $\rho$ ,

$$E[\rho(\gamma\|X - g\|_{\mathbb{H}} + (1 - \gamma)\|X - h\|_{\mathbb{H}})] \leq E[\gamma\rho(\|X - g\|_{\mathbb{H}}) + (1 - \gamma)\rho(\|X - h\|_{\mathbb{H}})],$$

and by the linearity of the expectation of a random variable,

$$\begin{aligned} E[\gamma\rho(\|X - g\|_{\mathbb{H}}) + (1 - \gamma)\rho(\|X - h\|_{\mathbb{H}})] &= \gamma E[\rho(\|X - g\|_{\mathbb{H}})] + (1 - \gamma)E[\rho(\|X - h\|_{\mathbb{H}})] \\ &= \gamma J(g) + (1 - \gamma)J(h). \end{aligned}$$

An analogous reasoning in the sample case proves the strict convexity of  $J_n$ .  $\square$

Note that in case  $\rho$  is convex and strictly increasing,  $P$  is not concentrated on a line in  $\mathbb{H}$  and  $X_1, \dots, X_n$  are almost surely not collinear, the population uniqueness and also the almost sure sample uniqueness are guaranteed by [Proposition 2.2](#).

Given that the minimization problem in [\(2\)](#) does not have an explicit solution in general, the representation in [Theorem 2.1](#) makes it possible to approximate functional location M-estimates by using a standard iteratively re-weighted least squares algorithm (see e.g. [Huber, P.J., 1964](#)). The algorithm can be summarized in the following steps (see also [Kim, J.S. and Scott, C.D., 2012](#)). Note that if  $\phi$  is non-increasing, then under assumptions C.1-C.4 it holds that  $\{g_{(k)}^M\}_{k=1}^{\infty}$  converges to  $\widehat{g}^M(X_1, \dots, X_n)$  in the norm  $\|\cdot\|_{\mathbb{H}}$ , as proven in [Kim, J.S. and Scott, C.D. \(2012\)](#).

*Step 1.* Select initial weights  $u_i^{(0)} \in \mathbb{R}$ , for  $i \in \{1, \dots, n\}$ , such that  $u_i^{(0)} \geq 0$  and  $\sum_{i=1}^n u_i^{(0)} = 1$ .

*Step 2.* Generate a sequence  $\{g_{(k)}^M\}_{k \in \mathbb{N}}$  by iterating the following procedure:

$$g_{(k)}^M = \sum_{i=1}^n u_i^{(k-1)} X_i, \quad u_i^{(k)} = \frac{\phi(\|X_i - g_{(k)}^M\|_{\mathbb{H}})}{\sum_{j=1}^n \phi(\|X_j - g_{(k)}^M\|_{\mathbb{H}})}.$$

Step 3. Terminate the algorithm when

$$\frac{\left| J_n \left( g_{(k+1)}^M \right) - J_n \left( g_{(k)}^M \right) \right|}{J_n \left( g_{(k)}^M \right)} < \varepsilon,$$

for some desired tolerance  $\varepsilon > 0$ .

In practice, under the conditions of the Representer Theorem the M-estimate of location of a sample of observations from a Hilbert space valued random element  $X$  can be approximated by this algorithm. Hence, an approximation is obtained even if the sufficient conditions for the measurability of the M-estimator of location given in Proposition 3.3 below (i.e. the Hilbert space is  $\sigma$ -compact and separable and the probability space is Polish) do not hold.

### 3. Properties of M-estimators of location for functional data

The loss  $\rho(\|h - g\|_{\mathbb{H}})$  is typically interpreted as a measure of the error when approximating the functional value  $h$  by means of the functional value  $g$ . Therefore, a sample M-estimator of location can be directly interpreted as a measure of the center of the (induced) probability distribution associated with the Hilbert space valued random variable. We now discuss properties of this M-estimator of location. We start with some equivariance properties.

**Proposition 3.1** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a Hilbert space and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable. Moreover, let  $\rho$  be a loss function satisfying C.1-C.3 and  $h_0 \in \mathbb{H}$ . If  $g^M(X)$  is an M-location value of  $X$ , then  $g^M(X) + h_0$  is an M-location value of  $X + h_0$ . In particular, if the M-location value  $g^M(X)$  is unique, then it is translation (i.e., shift) equivariant.*

*Proof.* Indeed, since  $\{h + h_0 : h \in \mathbb{H}\} = \mathbb{H}$  and due to the translational invariance of the norm  $\|\cdot\|_{\mathbb{H}}$ , one can conclude that

$$\begin{aligned} E [\rho(\|(X + h_0) - (g^M(X) + h_0)\|_{\mathbb{H}})] &= E [\rho(\|X - g^M(X)\|_{\mathbb{H}})] \\ &= \min_{h \in \mathbb{H}} E [\rho(\|X - h\|_{\mathbb{H}})] = \min_{h \in \mathbb{H}} E [\rho(\|(X + h_0) - (h + h_0)\|_{\mathbb{H}})] \\ &= \min_{h \in \mathbb{H}} E [\rho(\|(X + h_0) - h\|_{\mathbb{H}})]. \quad \square \end{aligned}$$

**Remark 3.1** Unfortunately, the M-location value is not scale equivariant in general. A strong extra condition about  $\rho$  is necessary ( $\rho$  being a homomorphism of the multiplicative group of positive real numbers), which is only fulfilled if  $\rho$  is a power function. Consequently, the resulting location M-estimators may depend heavily on the considered measurement units, similarly as in the real case (see e.g. Maronna, R.A. et al., 2006). In

order to avoid this, the procedure to choose the tuning parameters in loss functions such as (3)-(5) should take the distribution of the distances into account. To fix them, we can first calculate an initial robust estimator of location. In particular, we use the impartial trimmed mean of Cuesta-Albertos, J.A. and Fraiman, R. (2007) as in Colubi, A. and González-Rodríguez, G. (2015) for this purpose. Then, the tuning constants in the loss function are selected from the distribution of the distances between the observations and this initial estimate. For example,  $a$ ,  $b$  and  $c$  in (3)-(5) could be taken to be the median, the 75<sup>th</sup> percentile and the 85<sup>th</sup> percentile of these distances.

For real-valued random variables with a symmetric distribution it is well-known that the mean and the median (if unique) are equal to the point of symmetry. This property is expected for any sensible measure of the center of a symmetric random variable. We extend this property to Hilbert space valued random variables using the notion of symmetry as introduced by Vakhania, N. et al. (1987) and Paulauskas, V. and Račkauskas, A. (1989), for example. Proposition 3.2 confirms that the M-location value satisfies the desirable properties of a measure of the center of a random variable.

**Proposition 3.2** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a Hilbert space,  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable and  $\rho$  a loss function satisfying C.1-C.3. If  $X - h$  is symmetrically distributed for some  $h \in \mathbb{H}$ , then  $J$  is symmetric with respect to  $h$ . Therefore, if  $g_0$  minimizes  $J$ , then  $2h - g_0$  minimizes  $J$  too. Obviously, if the M-location value  $g^M(X)$  is unique, then  $g^M(X) = h$ .*

*Proof.* Since the random elements  $X - h$  and  $h - X$  have the same distribution, for any  $g \in \mathbb{H}$ ,

$$J(h + g) = E[\rho(\|X - h - g\|_{\mathbb{H}})] = E[\rho(\|h - X - g\|_{\mathbb{H}})] = J(h - g).$$

Trivially, if the M-location value is unique,  $g^M(X) = 2h - g^M(X)$  and, finally,  $g^M(X) = h$ .  $\square$

Some extra conditions on the probability space and Hilbert space allow us to show that the M-estimators of location are Borel measurable mappings.

**Proposition 3.3** *Let  $(\Omega, \mathcal{A}, P)$  be a Polish probability space,  $\mathbb{H}$  be a  $\sigma$ -compact and separable Hilbert space with associated norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable. Moreover, let  $(X_1, \dots, X_n)$  be a simple random sample from  $X$  and  $\rho$  a loss function satisfying C.1. Then, the M-estimator of location is a Borel measurable function.*

*Proof.* For two complete separable metric spaces,  $X$  and  $Y$ , and a function  $f : D \subseteq X \times Y \rightarrow \mathbb{R}$ , Brown, L.D. and Purves, R. (1973) give sufficient conditions for the existence of a Borel measurable mapping  $\varphi$  such that  $f(x, \varphi(x)) = \inf_y f(x, y)$ . Let  $\text{proj}(D)$  denote the set of all first coordinates of elements of  $D$ , then these sufficient conditions are

- $f$  is a real-valued Borel measurable function defined on a Borel subset  $D$  of  $X \times Y$ ,
- For each  $x \in \text{proj}(D)$ , the section  $D_x = \{y \in Y : (x, y) \in D\}$  is  $\sigma$ -compact and  $f(x, \cdot)$  is lower semi-continuous with respect to the relative topology on  $D_x$ .



To verify the first of these conditions for our setting, note that  $X \equiv \Omega$  is a Polish space, so  $X$  is indeed a complete separable metric space. Moreover,  $Y \equiv \mathbb{H}$  and a Hilbert space is always complete with the metric associated with the norm,  $d_{\|\cdot\|_{\mathbb{H}}}$ . Since  $\mathbb{H}$  is assumed to be separable, it follows that  $Y$  is also a complete separable metric space.

Now define  $f : \Omega \times \mathbb{H} \rightarrow \mathbb{R}$  such that  $f(\omega, h) = \frac{1}{n} \sum_{i=1}^n \rho(\|X_i(\omega) - h\|_{\mathbb{H}})$ . To show that  $f$  is a real-valued measurable function, note that  $D = \Omega \times \mathbb{H}$  is a Borel set, since it is the universal set and thus it belongs to any  $\sigma$ -algebra. The measurability of  $f$  now follows from the following facts:

- All  $X_i : \Omega \rightarrow \mathbb{H}$  are Borel measurable by definition of Hilbert space valued random variables. Then, the mappings  $f_i : \Omega \times \mathbb{H} \rightarrow \mathbb{H}$  with  $f_i(\omega, h) = X_i(\omega)$  are obviously Borel measurable with respect to the  $\sigma$ -algebra on the product space  $\Omega \times \mathbb{H}$  and the Borel  $\sigma$ -algebra on  $\mathbb{H}$  generated by the topology induced by  $\|\cdot\|_{\mathbb{H}}$ .
- The mapping  $f_0 : \Omega \times \mathbb{H} \rightarrow \mathbb{H}$  such that  $f_0(\omega, h) = h$  is also Borel measurable following an analogous reasoning as for  $f_i$ , since the identity is measurable.
- The norm  $\|\cdot\|_{\mathbb{H}}$  is a continuous function and thus Borel measurable.
- $\rho$  is assumed to be continuous and thus Borel measurable.
- The addition, product and composition of Borel measurable functions are Borel measurable as well.

In order to verify the second condition, take into account that for each  $\omega \in \text{proj}(D) = \Omega$ , the section  $D_\omega = \mathbb{H}$  is  $\sigma$ -compact.

For each  $\omega \in \text{proj}(D) = \Omega$  consider the function

$$\begin{aligned} f(\omega, \cdot) : \mathbb{H} &\longrightarrow \mathbb{R} \\ h &\longmapsto f(\omega, h) = \frac{1}{n} \sum_{i=1}^n \rho(\|X_i(\omega) - h\|_{\mathbb{H}}), \end{aligned}$$

then this function is continuous (and thus also lower semi-continuous) with respect to the relative topology on  $D_\omega = \mathbb{H}$ .

For each  $\omega \in \Omega$ , let  $h^*$  be any element of  $\mathbb{H}$ . It is now sufficient to check that, given any sequence  $\{h_n\}_{n \in \mathbb{N}} \subset \mathbb{H}$  such that  $h_n \xrightarrow[n]{\|\cdot\|_{\mathbb{H}}} h^*$  (i.e.,  $\|h_n - h^*\|_{\mathbb{H}} \xrightarrow[n]{} 0$ ), it holds that  $\lim_n f(\omega, h_n) = f(\omega, h^*)$ . Indeed, for all  $i = 1, \dots, n$ , the triangular inequality of the norm yields, that is,

$$\|X_i(\omega) - h^*\|_{\mathbb{H}} - \|h_n - h^*\|_{\mathbb{H}} \leq \|X_i(\omega) - h_n\|_{\mathbb{H}} \leq \|X_i(\omega) - h^*\|_{\mathbb{H}} + \|h^* - h_n\|_{\mathbb{H}}.$$

Since  $\|h_n - h^*\|_{\mathbb{H}} \xrightarrow[n]{} 0$ , we have that  $\|X_i(\omega) - h_n\|_{\mathbb{H}} \xrightarrow[n]{} \|X_i(\omega) - h^*\|_{\mathbb{H}}$ . The continuity of  $\rho$  implies that  $\rho(\|X_i(\omega) - h_n\|_{\mathbb{H}}) \xrightarrow[n]{} \rho(\|X_i(\omega) - h^*\|_{\mathbb{H}})$  and, obviously, the result follows because the continuity is preserved by addition and product of continuous functions.

Finally, the result by [Brown, L.D. and Purves, R. \(1973\)](#) guarantees the measurability

of the following function:

$$\begin{aligned} \varphi : I &\longrightarrow \mathbb{H} \\ \omega &\longmapsto g_0 \text{ s.t. } f(\omega, g_0) = \inf_{h \in \mathbb{H}} f(\omega, h) \\ &= g_0 \text{ s.t. } \frac{1}{n} \sum_{i=1}^n \rho(\|X_i(\omega) - g_0\|_{\mathbb{H}}) = \inf_{h \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^n \rho(\|X_i(\omega) - h\|_{\mathbb{H}}), \end{aligned}$$

where  $I = \{\omega \in \Omega : \text{for some } g_0 \in \mathbb{H}, f(\omega, g_0) = \inf_{h \in \mathbb{H}} f(\omega, h)\}$ . That is to say,  $\varphi$  is the function that assigns to each  $\omega \in \Omega$  for which the M-estimate of location exists, the corresponding value of the M-estimate of location.  $\square$

A key property of estimators is their strong consistency. From Theorem 2.1 it follows that the location M-estimator of a random sample from a Hilbert space valued random variable can be expressed as a randomly weighted mean of Hilbert space valued random variables. Although some limit theorems exist for the complex problem of randomly weighted means where the weights depend on the random variables  $X_i$  (see e.g. Ordóñez-Cabrera, M. *et al.*, 2012), it is not straightforward to generalize these results to our setting. Vandermeulen, R.A. and Scott, C.D. (2013) have recently discussed weak  $L^1$  consistency of the M-estimators of a density function in Kim, J.S. and Scott, C.D. (2012), but these results do not seem directly extensible to general Hilbert space valued random variables. Therefore, our approach is based on the result by Huber, P.J. (1967), who derived sufficient conditions for strong consistency of M-estimators in general settings. In the next proposition, we state assumptions that allow us to prove the strong consistency irrespectively of the considered norm. However, note that these assumptions include the local compactness and separability of the Hilbert space, which limits the range of applicability of this result to the finite dimensional case.

**Theorem 3.4** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a locally compact and second-countable Hilbert space with norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable. Moreover, let  $(X_1, \dots, X_n)$  be a simple random sample from  $X$  and  $\rho$  a loss function fulfilling C.1-C.3 such that either*

- $\rho$  is subadditive and unbounded,
- $\rho$  has linear upper and lower bounds with the same slope,
- $\rho$  is bounded.

*If the M-location value  $g^M(X)$  exists and is unique, then the M-estimator of location  $\widehat{g}^M(X_1, \dots, X_n)$  is a strongly consistent estimator of  $g^M(X)$ . That is,  $\|\widehat{g}^M(X_1, \dots, X_n) - g^M(X)\|_{\mathbb{H}}$  converges to 0 almost surely as  $n \rightarrow \infty$ .*

*Proof.* We are going to check that the sufficient conditions for strong consistency established by Huber, P.J. (1967) are all fulfilled under any of the assumptions above. First, consider the case that  $\rho$  is subadditive and unbounded.

Let denote by  $\mathcal{X} \subseteq \mathbb{H}$  the sample space of  $X$  and let us define the real-valued function  $q$  by

$$q : \mathcal{X} \times \mathbb{H} \rightarrow \mathbb{R}, \quad (x, h) \longmapsto q(x, h) := \rho(\|x - h\|_{\mathbb{H}})$$

and consider,

$$T_n : \mathcal{X}^n \rightarrow \mathbb{H}, \quad (x_1, \dots, x_n) \mapsto T_n(x_1, \dots, x_n) := \widehat{g^M}(x_1, \dots, x_n).$$

Then, by definition of the M-estimator of location, the sequence  $\{T_n\}_{n \in \mathbb{N}}$  satisfies, almost surely, that

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{i=1}^n q(x_i, T_n(x_1, \dots, x_n)) - \inf_{h \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^n q(x_i, h) \right]$$

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{i=1}^n \rho(\|x_i - T_n(x_1, \dots, x_n)\|_{\mathbb{H}}) - \inf_{h \in \mathbb{H}} \frac{1}{n} \sum_{i=1}^n \rho(\|x_i - h\|_{\mathbb{H}}) \right] = 0.$$

We now recall each of Huber's conditions and show that they are satisfied.

*Condition (H-1):* For each fixed  $h_0 \in \mathbb{H}$ , the function  $q_{h_0} : \mathcal{X} \rightarrow \mathbb{R}$ ,  $q_{h_0}(x) = q(x, h_0)$ , is  $\mathcal{B}_{\|\cdot\|_{\mathbb{H}}}$ -measurable and separable in Doob's sense (i.e., there is a  $P$ -null set  $N$  and a countable subset  $S \subset \mathbb{H}$  such that for every open set  $U \subset \mathbb{H}$  and every closed interval  $A$ , the sets  $V_1 = \{x \in \mathcal{X} : q(x, h) \in A \text{ for all } h \in U\}$  and  $V_2 = \{x \in \mathcal{X} : q(x, h) \in A \text{ for all } h \in U \cap S\}$  differ by at most a subset of  $N$ ).

First note that for an arbitrarily fixed  $h_0 \in \mathbb{H}$ , the function  $q_{h_0} : \mathcal{X} \rightarrow \mathbb{R}$  such that  $q_{h_0}(x) = q(x, h_0) = \rho(\|x - h_0\|_{\mathbb{H}})$  is  $\mathcal{B}_{\|\cdot\|_{\mathbb{H}}}$ -measurable because both the norm and  $\rho$  are continuous. The function  $q_{h_0}$  is also separable in Doob's sense because  $\mathbb{H}$  is second-countable, what implies the separability. Moreover, it implies that  $\mathbb{H}$  contains a countable dense subset, which we denote by  $S$ . Then, for every open set  $U \subset \mathbb{H}$  and every closed interval  $A$ , one can verify by *reductio ad absurdum* that the sets  $V_1 = \{x \in \mathcal{X} : q(x, h) \in A \text{ for all } h \in U\}$  and  $V_2 = \{x \in \mathcal{X} : q(x, h) \in A \text{ for all } h \in U \cap S\}$  coincide.

*Condition (H-2):* The function  $q$  is a.s. lower semi-continuous in  $h_0$ , that is to say,  $\inf_{h \in U_n} q(x, h) \rightarrow q(x, h_0)$  when the neighborhood  $U_n$  of  $h_0$  shrinks to  $\{h_0\}$ .

Indeed it will be proven that  $q$  is continuous in  $h_0$ . Given any arbitrary  $x \in \mathcal{X}$ , the function  $q_x : \mathbb{H} \rightarrow \mathbb{R}$  such that  $h \mapsto q_x(h) = \rho(\|x - h\|_{\mathbb{H}})$  is obviously continuous, since both  $\rho$  and  $\|\cdot\|_{\mathbb{H}}$  are continuous. Therefore,  $q_x(h) \xrightarrow{h \rightarrow h_0} q_x(h_0)$ .

*Condition (H-3):* There is a measurable function  $a : \mathcal{X} \rightarrow \mathbb{R}$  such that

$$E_{P_X}[q(\cdot, h) - a(\cdot)]^- < \infty \quad \text{for all } h \in \mathbb{H},$$

$$E_{P_X}[q(\cdot, h) - a(\cdot)]^+ < \infty \quad \text{for some } h \in \mathbb{H},$$

where  $P_X$  denotes the probability induced by  $X$ . Thus,  $\gamma(h) = E[q(\cdot, h) - a(\cdot)]$  is well-defined for all  $h$ .

Set  $a(x) = \rho(\|x\|_{\mathbb{H}})$ , which is a measurable function, and take any (fixed)  $h \in \mathbb{H}$ . By

the triangular inequality of  $\|\cdot\|_{\mathbb{H}}$  and the subadditivity of  $\rho$ , it holds that

$$\begin{aligned} E_{P_X} [\rho(\|x - h\|_{\mathbb{H}}) - a(x)]^- &= E_{P_X} [-\min\{\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}}), 0\}] \\ &= \int_{\{\rho(\|x\|_{\mathbb{H}}) > \rho(\|x - h\|_{\mathbb{H}})\}} [\rho(\|x\|_{\mathbb{H}}) - \rho(\|x - h\|_{\mathbb{H}})] dP_X \\ &\leq \rho(\|h\|_{\mathbb{H}}) \cdot P_X(\rho(\|x\|_{\mathbb{H}}) > \rho(\|x - h\|_{\mathbb{H}})) \\ &\leq \rho(\|h\|_{\mathbb{H}}) < \infty. \end{aligned}$$

Analogously,

$$\begin{aligned} E_{P_X} [\rho(\|x - h\|_{\mathbb{H}}) - a(x)]^+ &= E_{P_X} [\max\{\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}}), 0\}] \\ &= \int_{\{\rho(\|x\|_{\mathbb{H}}) \leq \rho(\|x - h\|_{\mathbb{H}})\}} [\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})] dP_X \\ &\leq \rho(\|h\|_{\mathbb{H}}) \cdot P_X(\rho(\|x\|_{\mathbb{H}}) \leq \rho(\|x - h\|_{\mathbb{H}})) \\ &\leq \rho(\|h\|_{\mathbb{H}}) < \infty. \end{aligned}$$

*Condition (H-4):* There is a  $g_0 \in \mathbb{H}$  such that  $\gamma(h) > \gamma(g_0)$  for all  $h \neq g_0$ .

Under the assumptions in this proposition, the M-location value exists and is unique, so that

$$E_{P_X} [\rho(\|x - g^M(X)\|_{\mathbb{H}})] = \min_{h \in \mathbb{H}} E_{P_X} [\rho(\|x - h\|_{\mathbb{H}})].$$

Thus,

$$\begin{aligned} g^M(X) &= \arg \min_{h \in \mathbb{H}} E_{P_X} [\rho(\|x - h\|_{\mathbb{H}})] \\ &= \arg \min_{h \in \mathbb{H}} \left( E_{P_X} [\rho(\|x - h\|_{\mathbb{H}})] - E_{P_X} [\rho(\|x\|_{\mathbb{H}})] \right) \\ &= \arg \min_{h \in \mathbb{H}} E_{P_X} [\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})] = \arg \min_{h \in \mathbb{H}} \gamma(h), \end{aligned}$$

so  $g_0 = g^M(X)$  fulfills this assumption.

*Condition (H-5):* There exists a continuous function  $b(h) > 0$  such that

- for some integrable  $s$ ,  $\inf_{h \in \mathbb{H}} [q(x, h) - a(x)]/b(h) \geq s(x)$ ;
- $\liminf_{h \rightarrow \infty} b(h) > \gamma(g_0)$ ; where  $\infty$  denotes the point at infinity in its one-point compactification;
- $E_{P_X} [\liminf_{h \rightarrow \infty} [q(\cdot, h) - a(\cdot)]/b(h)] \geq 1$ .

Consider the continuous function  $b : \mathbb{H} \rightarrow (0, \infty)$  given by  $b(h) = \rho(\|h\|_{\mathbb{H}}) + 1$ . We check that the three requirements hold for this function under the assumptions in this proposition.

- Using the properties of  $\rho$  we obtain that

$$\inf_{h \in \mathbb{H}} \frac{\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h\|_{\mathbb{H}}) + 1} \geq \inf_{h \in \mathbb{H}} \frac{-\rho(\|h\|_{\mathbb{H}})}{\rho(\|h\|_{\mathbb{H}}) + 1} \geq -1.$$

Hence, the first requirement holds for the integrable function  $s(x) := -1$ .

- Similarly, we have that

$$\liminf_{h \rightarrow \infty} b(h) > \gamma(g_0) = E_{P_X} [\rho(\|x - g_0\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})].$$

- We now check that in case  $\rho$  is subadditive and unbounded, it holds that

$$E_{P_X} \left[ \liminf_{h \rightarrow \infty} \frac{\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{b(h)} \right] \geq 1.$$

Consider an arbitrary sequence  $(h_n)_{n \in \mathbb{N}} \rightarrow \infty$ . Note that for any  $x \in \mathcal{X}$  the sequence

$$\left\{ \inf_{k \geq n} \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \right\}_n$$

is monotonically increasing and is bounded above by 1, since for all  $k \in \mathbb{N}$  applying the triangular inequality and the subadditivity of  $\rho$  yields

$$\frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \leq \frac{\rho(\|h_k\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \leq 1.$$

Hence, the sequence converges to its supremum

$$\begin{aligned} \liminf_{h \rightarrow \infty} \frac{\rho(\|x - h\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h\|_{\mathbb{H}}) + 1} &= \lim_{n \rightarrow \infty} \left( \inf_{k \geq n} \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \right) \\ &= \sup_n \left( \inf_{k \geq n} \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \right). \end{aligned}$$

We now show that this supremum is at least equal to 1. By *reductio ad absurdum*, suppose that

$$\sup_n \left( \inf_{k \geq n} \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \right) \leq 1 - \varepsilon,$$

for some  $\varepsilon > 0$ . This yields a contradiction because we can find an index  $n^* \in \mathbb{N}$  such that

$$\inf_{k \geq n^*} \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} > 1 - \varepsilon$$

because for all  $k \geq n^*$

$$\frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \geq 1 - \frac{\varepsilon}{2} > 1 - \varepsilon.$$

To find  $n^*$ , set  $M = \frac{2}{\varepsilon} - 1 + \frac{4}{\varepsilon} \cdot \rho(\|x\|_{\mathbb{H}}) \in \mathbb{R}$ . Since  $\rho$  is unbounded, we can then select  $M^* > 0$  such that  $\rho(M^*) > M$ . Recall that  $\lim_{n \rightarrow \infty} h_n = \infty$ , so there exists a  $n^* \in \mathbb{N}$  such that for all  $n \geq n^*$ ,  $\|h_n\|_{\mathbb{H}} > M^*$ . Therefore,

$$\rho(\|h_n - x\|_{\mathbb{H}}) \geq \rho(\|h_n\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}}) \geq \rho(M^*) - \rho(\|x\|_{\mathbb{H}}) > M - \rho(\|x\|_{\mathbb{H}}),$$

because  $\rho$  is subadditive and non-decreasing. It is now easy to check that  $1 - \varepsilon/2$  is a lower bound of the sequence

$$\left\{ \frac{\rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}})}{\rho(\|h_k\|_{\mathbb{H}}) + 1} \right\}_{k \geq n^*}.$$

Indeed, for any  $k \geq n^*$  we have that,

$$\begin{aligned} \rho(\|x - h_k\|_{\mathbb{H}}) - \rho(\|x\|_{\mathbb{H}}) &= \left(1 - \frac{\varepsilon}{2}\right) \rho(\|x - h_k\|_{\mathbb{H}}) + \frac{\varepsilon}{2} \rho(\|x - h_k\|_{\mathbb{H}}) \\ &- \left(1 - \frac{\varepsilon}{2}\right) \rho(\|x\|_{\mathbb{H}}) - \frac{\varepsilon}{2} \rho(\|x\|_{\mathbb{H}}) \geq \left(1 - \frac{\varepsilon}{2}\right) \rho(\|h_k\|_{\mathbb{H}}) - \left(1 - \frac{\varepsilon}{2}\right) \rho(\|x\|_{\mathbb{H}}) \\ &\quad + \frac{\varepsilon}{2} \rho(\|x - h_k\|_{\mathbb{H}}) - \left(1 - \frac{\varepsilon}{2}\right) \rho(\|x\|_{\mathbb{H}}) - \frac{\varepsilon}{2} \rho(\|x\|_{\mathbb{H}}) \\ &= \left(1 - \frac{\varepsilon}{2}\right) \rho(\|h_k\|_{\mathbb{H}}) + \frac{\varepsilon}{2} \rho(\|x - h_k\|_{\mathbb{H}}) - \left(2 - \frac{\varepsilon}{2}\right) \rho(\|x\|_{\mathbb{H}}) \\ &> \left(1 - \frac{\varepsilon}{2}\right) \rho(\|h_k\|_{\mathbb{H}}) + \frac{\varepsilon}{2} \left(\frac{2}{\varepsilon} - 1 + \left(\frac{4}{\varepsilon} - 1\right) \rho(\|x\|_{\mathbb{H}})\right) - \left(2 - \frac{\varepsilon}{2}\right) \rho(\|x\|_{\mathbb{H}}) \\ &= \left(1 - \frac{\varepsilon}{2}\right) \rho(\|h_k\|_{\mathbb{H}}) + 1 - \frac{\varepsilon}{2} = \left(1 - \frac{\varepsilon}{2}\right) (\rho(\|h_k\|_{\mathbb{H}}) + 1), \end{aligned}$$

which shows the third requirement and hence the strong consistency in the case that  $\rho$  is subadditive and unbounded.

When  $\rho$  has linear upper and lower bounds sharing the slope (like it happens with Huber's loss function), it is possible to follow a similar reasoning as above, taking into account the bounds  $m x - n_1 \leq \rho(x) \leq m x + n_2$ , being  $m > 0, n_1, n_2 \geq 0$ , for all  $x \geq 0$ .

When  $\rho$  is bounded (like the Tukey biweight or Hampel loss functions), such an upper bound allows us to easily check conditions (H-1) to (H-5)(ii), but not condition (H-5)(iii). To complete the proof, it is necessary to show that there is a compact set  $C$  in the parameter space such that the sequence of M-estimators of location almost surely ultimately stays in  $C$ , as commented in [Huber, P.J. \(1967\)](#). Thanks to the parameter space being locally compact,  $g_0$  has a compact neighborhood, which we will denote  $C$ . Note that for any element  $h^* \notin C$  we can fix  $\varepsilon = \gamma(h^*) - \gamma(g_0) > 0$ . Using the strong law of large numbers, there exists  $n_0 \in \mathbb{N}$  such that for all  $n \geq n_0$

$$\begin{aligned} \inf_h \frac{1}{n} \sum_{i=1}^n \rho(\|x_i - h\|_{\mathbb{H}}) - \rho(\|x_i\|_{\mathbb{H}}) &\leq \frac{1}{n} \sum_{i=1}^n \rho(\|x_i - g_0\|_{\mathbb{H}}) - \rho(\|x_i\|_{\mathbb{H}}) \\ &\leq \gamma(g_0) + \frac{\varepsilon}{4} < \gamma(g_0) + \frac{\varepsilon}{2} = \gamma(h^*) - \frac{\varepsilon}{2} \leq \frac{1}{n} \sum_{i=1}^n \rho(\|x_i - h^*\|_{\mathbb{H}}) - \rho(\|x_i\|_{\mathbb{H}}) \quad \square \end{aligned}$$

We now investigate the robustness of functional location M-estimators by means of their finite sample breakdown point. As indicated by [Cuevas, A. et al. \(2007\)](#), this measure of robustness, originally introduced by [Hampel, F.R. \(1971\)](#) and further formalized

by [Donoho, D.L. and Huber, P.J. \(1983\)](#), can be adapted to estimators taking values in general metric spaces. The value of the finite sample breakdown point corresponds to the minimum proportion of observations in the sample that need to be perturbed to make the distance between the estimates based on the original and contaminated samples arbitrarily large. The following result shows that functional location M-estimators can have a high finite sample breakdown point of approximately 0.5, which means that they give a reliable performance even when almost half of the observations are contaminated.

**Theorem 3.5** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space, let  $\mathbb{H}$  be a Hilbert space with norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable. Moreover, let  $(x_1, \dots, x_n)$  be a sample obtained from  $X$  and  $\rho$  a loss function fulfilling C.1-C.3 such that the corresponding M-estimator of location exists and is unique. Then, the finite sample breakdown point of the M-estimator of location is at most  $\frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$ , where  $\lfloor \cdot \rfloor$  denotes the floor function. Under any of the additional assumptions:*

- $\rho$  admits linear upper and lower bounds with the same slope,
- $\rho$  has a finite upper bound  $C^*$  and satisfies

$$\rho \left( \max_{1 \leq i, j \leq n} \|x_i - x_j\|_{\mathbb{H}} \right) < \frac{n - 2 \lfloor \frac{n-1}{2} \rfloor}{n - \lfloor \frac{n-1}{2} \rfloor - 1} \cdot C^*; \quad (6)$$

the finite sample breakdown point is exactly  $\frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$ .

*Proof.* The proof of the upper bound  $\frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$  is an extension of the analogous result in the real setting, due to the translational equivariance of the M-estimators of location. When the loss function has linear upper and lower bounds with the same slope, it is possible to follow a reasoning similar as in [Lopuhaä, H.P. and Rousseeuw, P.J. \(1991\)](#) to guarantee that the bound  $\frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$  is indeed attained.

Let us now consider the case of a loss function  $\rho$  with finite upper bound. When condition (6) holds, then it can be shown that the finite sample breakdown point is exactly  $\frac{1}{n} \lfloor \frac{n+1}{2} \rfloor$ . Indeed, we prove that the M-estimate always belongs to a ball  $B$  of finite radius (only depending on the original sample) when at most  $\lfloor \frac{n-1}{2} \rfloor$  observations from the original sample  $(x_1, \dots, x_n)$  are perturbed. Let  $\mathbf{Y}_{\mathbf{n}, \mathbf{k}} = (y_1, \dots, y_n)$  denote the perturbed sample with  $k \leq \lfloor \frac{n-1}{2} \rfloor$  modified observations and  $I_{\mathbf{Y}_{\mathbf{n}, \mathbf{k}}} \subseteq \{1, \dots, n\}$  the subset of indices representing the original observations that belong to  $\mathbf{Y}_{\mathbf{n}, \mathbf{k}}$ . Define  $B$  as the ball centered at  $x_{i_0}$ , where  $i_0 = \min_{i \in I_{\mathbf{Y}_{\mathbf{n}, \mathbf{k}}}} \max_{j \in I_{\mathbf{Y}_{\mathbf{n}, \mathbf{k}}}} \|x_i - x_j\|_{\mathbb{H}}$ , with radius  $R := \max_{1 \leq j \leq n} \|x_{i_0} - x_j\|_{\mathbb{H}} + \min\{t > 0; \rho(t) = C^*\}$ .

By *reductio ad absurdum*, let us suppose that the M-estimate  $\widehat{g^M}(\mathbf{Y}_{\mathbf{n}, \mathbf{k}})$  does not belong to  $B$ . Then,

$$\sum_{i=1}^n \rho(\|y_i - \widehat{g^M}(\mathbf{Y}_{\mathbf{n}, \mathbf{k}})\|_{\mathbb{H}}) \geq \sum_{i \in I_{\mathbf{Y}_{\mathbf{n}, \mathbf{k}}}} \rho(\|x_i - \widehat{g^M}(\mathbf{Y}_{\mathbf{n}, \mathbf{k}})\|_{\mathbb{H}}) \geq (n - k) C^*.$$

On the other hand,

$$\sum_{i=1}^n \rho(\|y_i - x_{i_0}\|_{\mathbb{H}}) \leq \sum_{i \in I_{\mathbf{Y}_{n,k}}} \rho(\|x_i - x_{i_0}\|_{\mathbb{H}}) + k C^*.$$

Since  $\rho$  vanishes at 0 and is non-decreasing,

$$\begin{aligned} \sum_{i \in I_{\mathbf{Y}_{n,k}}} \rho(\|x_i - x_{i_0}\|_{\mathbb{H}}) &= \sum_{\substack{i \in I_{\mathbf{Y}_{n,k}} \\ i \neq i_0}} \rho(\|x_i - x_{i_0}\|_{\mathbb{H}}) \leq (n - k - 1) \rho\left(\max_{1 \leq i, j \leq n} \|x_i - x_j\|_{\mathbb{H}}\right) \\ &< (n - k - 1) \cdot \frac{n - 2 \lfloor \frac{n-1}{2} \rfloor}{n - \lfloor \frac{n-1}{2} \rfloor - 1} \cdot C^* \leq (n - k - 1) \cdot \frac{n - 2k}{n - k - 1} \cdot C^* = (n - 2k) C^*. \end{aligned}$$

Finally,

$$\begin{aligned} \sum_{i=1}^n \rho(\|y_i - x_{i_0}\|_{\mathbb{H}}) &\leq \sum_{i \in I_{\mathbf{Y}_{n,k}}} \rho(\|x_i - x_{i_0}\|_{\mathbb{H}}) + k C^* \\ &< (n - k) C^* \leq \sum_{i=1}^n \rho(\|y_i - \widehat{g}^M(\mathbf{Y}_{n,k})\|_{\mathbb{H}}), \end{aligned}$$

which is a contradiction.  $\square$

Note that for classes of loss functions such as the Tukey or Hampel class, condition (6) translates into a condition on the tuning parameters. It states that the tuning parameters should be chosen such that the loss function is well adapted to the data.

The finite sample breakdown point is a global measure of robustness of an estimator. On the other hand, the influence function is a well-known measure of the local robustness of an estimator. The definition for the influence function of a functional  $T$  at a distribution  $F$  is given by

$$IF(x'; T, F) = \lim_{s \rightarrow 0} \frac{T((1-s)F + s\delta_{x'}) - T(F)}{s},$$

where  $\delta_{x'}$  represents a discrete distribution that assigns probability 1 to the point  $x'$ . Hence,  $IF(x'; T, F)$  measures the change of the estimator  $T$  when the distribution  $F$  is contaminated with infinitesimal probability mass at  $x'$ . This definition of the influence function can also be used for Hilbert space valued statistics. Let  $F_s$  denote  $(1-s)F + s\delta_{x'}$  for any fixed  $x' \in \mathbb{H}$ . For the M-estimators of functional location the influence function can be obtained by adapting the results of [Kim, J.S. and Scott, C.D. \(2012\)](#) as in the following two theorems.

**Theorem 3.6** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a Hilbert space with norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable such that  $\|X\|_{\mathbb{H}}$  is bounded. Moreover, let  $\rho$  be a loss function satisfying C.1-C.3 such that  $\phi$  is Lipschitz continuous. We assume that  $g_{F_s}^M \xrightarrow{s \rightarrow 0} g_F^M$ . If  $\dot{g}_F^M := \lim_{s \rightarrow 0} (g_{F_s}^M - g_F^M)/s$  exists, then*

$$IF(x'; g^M, F) = \dot{g}_F^M,$$



where  $\dot{g}_F^M \in \mathbb{H}$  satisfies

$$\begin{aligned} \dot{g}_F^M \int \phi(\|X - g_F^M\|_{\mathbb{H}}) dF + \int \frac{\langle \dot{g}_F^M, X - g_F^M \rangle_{\mathbb{H}}}{\|X - g_F^M\|_{\mathbb{H}}^3} \cdot q(\|X - g_F^M\|_{\mathbb{H}})(X - g_F^M) dF \\ = (x' - g_F^M)\phi(\|x' - g_F^M\|_{\mathbb{H}}), \end{aligned}$$

and  $q(x) = x\rho''(x) - \rho'(x)$ .

Given any  $x' \in \mathbb{H}$ , we can define the following functions, for  $s \in (0, 1]$ ,

$$\begin{aligned} J_s : \mathbb{H} &\longrightarrow \mathbb{R} \\ h &\longmapsto J_s(h) = (1-s)E[\rho(\|X - h\|_{\mathbb{H}})] + s\rho(\|x' - h\|_{\mathbb{H}}). \end{aligned}$$

The weakest notion of convergence that would guarantee that minimizers of  $J_s$  converge to a minimizer of  $J$  is the epi-convergence or  $\Gamma$ -convergence (see e.g. [Rockafellar, R.T. and Wets, R.J.B., 1998](#); [Jerrard, R.L. and Sternberg, P., 2009](#)). Therefore, if the sequence of functionals  $J_s$  epi-converges to  $J$ ,  $g_F^M$  is unique and the sequence  $\{g_{F_s}^M\}_{s \rightarrow 0}$  is convergent,  $g_{F_s}^M \xrightarrow{s \rightarrow 0} g_F^M$  holds.

When the general distribution  $F$  is replaced by the empirical  $F_n$ , then it is possible to find  $\dot{g}_{F_n}^M$  explicitly.

**Theorem 3.7** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $\mathbb{H}$  a Hilbert space with norm  $\|\cdot\|_{\mathbb{H}}$  and  $X : \Omega \rightarrow \mathbb{H}$  an associated Hilbert space valued random variable such that  $\|X\|_{\mathbb{H}}$  is bounded. Moreover, let  $(x_1, \dots, x_n)$  be a sample of independent observations obtained from  $X$  and  $\rho$  a loss function satisfying C.1-C.3 such that  $\phi$  is Lipschitz continuous. If  $g_{F_n, s}^M \xrightarrow{s \rightarrow 0} g_{F_n}^M$  (a sufficient condition is that  $J_n$  is strictly convex) and the matrix  $K' = (\langle z_i, z_j \rangle_{\mathbb{H}})_{i=1, j=1}^{n+1}$  is positive definite, where  $z_i = x_i$  for  $i = 1 \dots, n$  and  $z_{n+1} = x'$ , then*

$$IF(x'; g^M, F_n) = \sum_{i=1}^n \alpha_i x_i + \alpha' x'.$$

Using the notation  $\mathbf{k}' = (\langle x', x_i \rangle_{\mathbb{H}})_{i=1}^n$ ,  $\gamma = \sum_{i=1}^n \phi(\|x_i - g_{F_n}^M\|_{\mathbb{H}})$ , and denoting by  $Q$  the diagonal matrix with  $Q_{ii} = \frac{q(\|x_i - g_{F_n}^M\|_{\mathbb{H}})}{\|x_i - g_{F_n}^M\|_{\mathbb{H}}^3}$ ,  $I_n$  the  $n \times n$  identity matrix and  $\mathbf{u}$  the vector of observation weights for the M-estimate as defined in the Representer Theorem, we have that  $\alpha' = n\phi(\|x' - g_{F_n}^M\|_{\mathbb{H}})/\gamma$  and  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)^t$  is the solution of the following system of linear equations

$$\begin{aligned} [\gamma I_n + (I_n - \mathbf{1}\mathbf{u}^t)^t Q (I_n - \mathbf{1}\mathbf{u}^t) K] \boldsymbol{\alpha} \\ = -n\phi(\|x' - g_{F_n}^M\|_{\mathbb{H}}) \mathbf{u} - \boldsymbol{\alpha}^t (I_n - \mathbf{1}\mathbf{u}^t)^t Q (I_n - \mathbf{1}\mathbf{u}^t) \mathbf{k}', \end{aligned}$$

with  $K = (\langle x_i, x_j \rangle_{\mathbb{H}})_{i, j=1}^n$ .

Notice that  $\alpha'$  represents how the M-estimator changes when an infinitesimal amount of contamination is introduced at  $x'$ . Since  $\phi(t) = \rho'(t)/t$ , for robust loss functions the effect of the contamination reduces quickly when  $x'$  lies further away from the observations in the sample.

## 4. Finite-sample behavior

In this section, we present the results of a simulation study to investigate the finite-sample behavior of the functional M-estimators of location. First, in Subsection 4.1, we numerically investigate the finite sample breakdown point of functional M-estimators of location. With a simulation study we then compare M-estimators of location with trimmed means for functional data. These results are presented in Subsection 4.2.

### 4.1. Finite sample breakdown point

For this simulation study, we consider the Hilbert space  $\mathbb{H} = L^2([0, 1])$ , the space of square Lebesgue integrable functions defined on the interval  $[0, 1]$ . Given any  $f \in L^2([0, 1])$ , the associated norm is  $\|f\|_{\mathbb{H}} = \|f\|_2 = \sqrt{\int_{[0,1]} f^2 d\lambda}$ , where  $\lambda$  represents the Lebesgue measure on  $[0, 1]$ . The curves are represented by their values at  $T = 30$  equidistant points in the domain  $[0, 1]$ . To empirically illustrate the behavior of functional M-estimates of location under data contamination, a sample of  $n$  functions,  $\{X_i\}_{i=1}^n$ , has been generated from the following model:

$$X_i(t) = 4t + e_i(t), \quad i = 1, \dots, n,$$

where  $e_i(t)$  is a Gaussian stochastic process with mean 0 and covariance function  $\gamma(s, t) = e^{-|t-s|}$ . Two sample sizes  $n$  will represent the even ( $n = 100$ ) and odd ( $n = 101$ ) cases. In each situation, using as starting value the impartial trimmed mean for functional data (see Cuesta-Albertos, J.A. and Fraiman, R., 2007; Colubi, A. and González-Rodríguez, G., 2015) with trimming proportion .5, the tuning parameter for the Huber loss function has been selected as commented in Remark 3.1 and the corresponding Huber M-estimate has been computed.

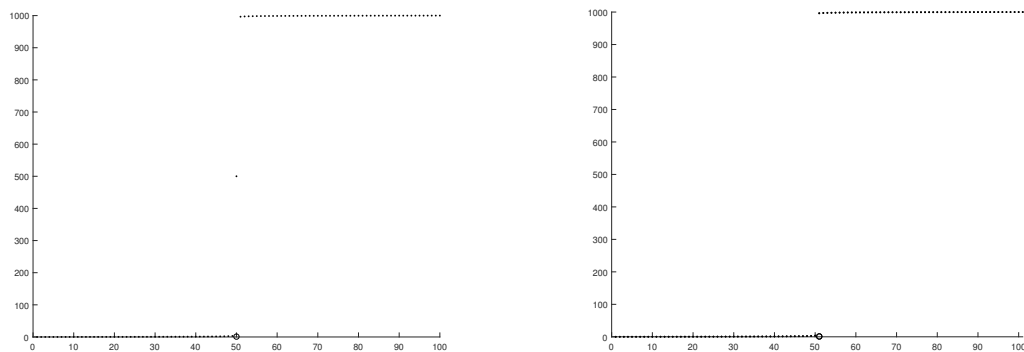
The impartial trimmed mean for functional data is defined as follows. Let  $(\Omega, \mathcal{A}, P)$  be a probability space,  $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$  be a Hilbert space and  $X$  be an  $\mathbb{H}$ -valued random variable. Consider a sample of independent observations  $(x_1, \dots, x_n)$  and a trimming level  $\alpha \in (0, 1)$ . Let  $\mathcal{E} = \{E \subset \{1, \dots, n\} : \#E = h\}$  be the collection of all the subsets of size  $n - \lfloor n\alpha \rfloor$ . Moreover, for any  $E \in \mathcal{E}$ , let  $\bar{x}_E$  denote the functional mean of the observations  $\{x_j : j \in E\}$ , then the impartial trimmed mean (ITM) is defined as

$$\hat{g}_{ITM, \alpha}(x_1, \dots, x_n) = \bar{x}_{\hat{E}} \quad \text{with} \quad \hat{E} = \operatorname{argmin}_{E \in \mathcal{E}} \frac{1}{n - \lfloor n\alpha \rfloor} \sum_{i \in E} \|x_i - \bar{x}_E\|_{\mathbb{H}}^2.$$

Cuesta-Albertos, J.A. and Fraiman, R. (2007) proposed an algorithm to calculate an approximation of the impartial trimmed mean. This approximation always corresponds to one of the observations in the sample. We use a more refined algorithm that is obtained by adapting the concentration algorithm of Rousseeuw, P.J. and Van Driessen, K. (2006) to our setting. Our adaptation is similar to the recent adaptation for the case of fuzzy-valued random variables in Colubi, A. and González-Rodríguez, G. (2015). By including

the sample observations in the set of starting values of our algorithm, we can guarantee that its solution is at least as good as the approximation with the simple algorithm of [Cuesta-Albertos, J.A. and Fraiman, R. \(2007\)](#).

Afterwards,  $i \in \{1, \dots, n\}$  observations have been highly contaminated (concretely,  $i$  functions have been translated  $10^3$  units) and Huber M-estimates have been computed again, as well as the distances between non-contaminated and contaminated estimates for each amount of modified observations, as shown in Figure 1. The value  $\circ$  represents the minimum number of perturbed observations that makes the distance between the non-contaminated and the contaminated corresponding M-estimates increase arbitrarily, i.e., the finite sample breakdown point. Similar plots can be obtained using different sample sizes, translations, tuning parameter and loss functions fulfilling the conditions in Theorem 3.5, such as Tukey and Hampel.



**Figure 1.** Empirical value ( $\circ$ ) obtained for the finite sample breakdown point of the M-estimators for functional data and distances between the non-contaminated and contaminated estimates using the Huber loss function when the sample size is even (left) and odd (right).

## 4.2. Comparative study

For the comparative study, we use the same designs as in previous works on robust functional data analysis (see e.g. [Fraiman, R. and Muniz, G., 2001](#); [López-Pintado, S. and Romo, J., 2009](#)). The first simulation model does not contain any contamination. The other models contain different types of outlying curves. Models 2-5 contain magnitude outliers, i.e. curves that lie far from the center. Models 6-9 contain shape outliers, which are curves that do not necessarily lie far from the center, but they show a shape or pattern that differs from the majority. A recent overview of the types of outliers that can occur in functional data can be found in [Hubert, M. et al. \(2015\)](#). For each of the simulation settings, we generate  $N = 500$  samples with sample size  $n = 50$  or  $n = 80$ . The contamination fraction  $\epsilon$  is either 5% or 10%.

For models 1-5, we consider the Hilbert space  $\mathbb{H} = L^2([0, 1])$ . Given any  $f \in L^2([0, 1])$ , the associated norm is  $\|f\|_{\mathbb{H}} = \|f\|_2$  and the curve is represented by its values at

$T = 30$  equidistant points in the domain  $[0, 1]$ . Therefore, the norm  $\|f\|_2$  of a curve  $f$  is approximated by using the available set of points through the expression  $\|f\|_2 \approx \sqrt{\frac{1}{T} \sum_{k=1}^T f(k)^2}$ . The tuning parameters for the Huber, Tukey and Hampel loss functions are chosen as explained in Remark 3.1.

Let  $\{\varepsilon_i\}_{i=1}^n$  and  $\{\sigma_i\}_{i=1}^n$  be two independent sequences of random variables following a Bernoulli (with parameter  $\epsilon$ ) and a discrete uniform on  $\{-1, 1\}$  distribution, respectively. Then, the considered models can be described as follows.

- *Model 1:* The curves  $X_i(t)$  are generated according to the following distribution:

$$X_i(t) = 4t + e_i(t), \quad i = 1, \dots, n,$$

where  $e_i(t)$  is a Gaussian stochastic process with mean 0 and covariance function  $\gamma(s, t) = e^{-|t-s|}$ .

- *Model 2:* Symmetric contamination obtained by generating the curves according to

$$Y_i(t) = X_i(t) + \varepsilon_i \sigma_i M, \quad 1 \leq i \leq n.$$

- *Model 3:* Asymmetric contamination given by the curves

$$Y_i(t) = X_i(t) + \varepsilon_i M, \quad 1 \leq i \leq n.$$

- *Model 4:* Partial (trajectories) contamination given by the curves:

$$Y_i(t) = \begin{cases} X_i(t) + \varepsilon_i \sigma_i M, & \text{if } t \geq T_i \\ X_i(t), & \text{if } t < T_i \end{cases}$$

for  $1 \leq i \leq n$  and where the corresponding  $T_i$  is a random number generated from a uniform distribution on  $(0, 1)$ .

- *Model 5:* Peaks contamination by generating the curves according to

$$Y_i(t) = \begin{cases} X_i(t) + \varepsilon_i \sigma_i M, & \text{if } T_i \leq t \leq T_i + l \\ X_i(t), & \text{if } t \notin [T_i, T_i + l] \end{cases}$$

for  $1 \leq i \leq n$ ,  $l = 2/30$  and the corresponding  $T_i$  being a random number generated from a uniform distribution on  $(0, 1 - l)$ .

The value of the contamination size constant  $M$  is either 5 or 25. In models 6-9 we consider shape outliers instead of magnitude outliers. Since shape outliers do not necessarily lie far from the center, the Euclidean distance corresponding to the  $L^2$  norm  $\|f\|_2$  is not appropriate to identify them. However, the deviating behavior of shape outliers is more easily picked up when taking derivatives of the curves into account (see e.g. Claeskens, G. et al., 2014). Therefore, in models 6-9 we consider the Sobolev space  $\mathbb{H} = W^{1,2}([0, 1]) = \{f \in L^2([0, 1]) : Df \in L^2([0, 1])\}$  with norm

$$\|f\|_{\mathbb{H}} = \|f\|_W = \sqrt{\|f\|_2^2 + \|Df\|_2^2},$$

and corresponding distance

$$d(f, g)_W = \sqrt{\|f - g\|_2^2 + \|D(f - g)\|_2^2}.$$

Based on this distance function, shape outliers can be identified better because it does not only involve the Euclidean distance between the functions, but also between their derivatives.

For the settings with shape outliers, the regular curves are generated according to

$$X_i(t) = g(t) + e_{1i}(t), \quad 1 \leq i \leq n,$$

where  $g(t)$  is given by either  $g(t) = 4t$  as in models 1-5 or  $g(t) = 4t^2$ . The errors  $e_{1i}(t)$  come from a Gaussian process with mean 0 and covariance function  $\gamma_1(s, t) = e^{-|t-s|^2}$ . The contaminated curves are generated according to

$$Y_i(t) = g(t) + e_{2i}(t), \quad 1 \leq i \leq n,$$

where the Gaussian process  $e_{2i}(t)$  still has mean 0, but its covariance function now equals  $\gamma_2(s, t) = k \cdot e^{-c|t-s|^\mu}$  with nonnegative parameters  $k$ ,  $c$  and  $\mu$ . More details about this family of models are given in [Wood, A.T.A. and Chan, G. \(1994\)](#). The parameters in the covariance function control the shape of the curves. Increasing  $\mu$  and  $k$  yields softer curves, while increasing  $c$  results in more irregular curves. Let  $B$  be a Bernoulli random variable with parameter  $\epsilon$ , then the data generating model is given by

$$Z_i(t) = (1 - B)X_i(t) + BY_i(t), \quad 1 \leq i \leq n.$$

The following choices for the parameters are considered:

- *Model 6:*  $k = 1$ ,  $c = 1$ ,  $\mu = .2$  and  $g(t) = 4t$
- *Model 7:*  $k = 1$ ,  $c = 1$ ,  $\mu = .1$  and  $g(t) = 4t$
- *Model 8:*  $k = 1$ ,  $c = 1$ ,  $\mu = .2$  and  $g(t) = 4t^2$
- *Model 9:*  $k = 1$ ,  $c = 1$ ,  $\mu = .1$  and  $g(t) = 4t^2$

Next to the sample mean of the functional data and the functional M-estimators of location based on the Huber, Tukey and Hampel loss functions, we also consider the impartial trimmed mean for functional data and the depth based trimmed mean of [Fraiman, R. and Muniz, G. \(2001\)](#), so we first recall the definition of the latter measure.

The depth based trimmed mean of [Fraiman, R. and Muniz, G. \(2001\)](#) is based on the sample depth of the observations, given by

$$D(x_i)_n = 1 - \frac{1}{T} \sum_{t=0}^T \left| \frac{1}{2} - \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{(-\infty, x_i(t)]}(x_j(t)) \right|, \quad i \in \{1, \dots, n\}.$$

The sample depth of the observations provides a center outward ordering of the curves. For a given trimming level  $\alpha \in (0, 1)$ , the depth based trimmed mean is given by the average of the  $n - \lfloor n\alpha \rfloor$  curves with largest sample depth. That is, the sample depth trimmed mean (DTM) equals

$$\hat{g}_{DTM, \alpha}(x_1, \dots, x_n) = \frac{\sum_{i=1}^{n - \lfloor n\alpha \rfloor} x_{i:n}}{n - \lfloor n\alpha \rfloor},$$

where  $x_{i:n}$  are the observations ordered according to decreasing sample depth. For both *ITM* and *DTM*, the trimming proportion has been chosen to be  $\alpha = 0.2$ , which is a common value (see, e.g., [Fraiman, R. and Muniz, G., 2001](#)) when no more information is available (in practice, we would not know that  $\epsilon$  is equal to 0.05 or 0.1). Obviously, the performance of the trimmed means could be improved by developing an optimal trimming proportion selection procedure.

**Table 1.** MSE and corresponding standard error (between brackets) of the mean, the trimmed mean (ITM), the depth trimmed mean (DTM), and Huber (M-Huber), Tukey (M-Tukey) and Hampel (M-Hampel) M-estimates in Models 1-5

$n$	$\epsilon$	$M$	Estimator	Model 1	Model 2	Model 3	Model 4	Model 5
50	.05	5	Mean	<b>.019203</b> (.019246)	.042175 (.056241)	.102664 (.1191556)	.031277 (.031111)	<b>.020245</b> (.019566)
			ITM	.041228 (.044861)	.036191 (.040044)	.040621 (.047352)	.042192 (.048873)	.040540 (.046823)
			DTM	.024925 (.025128)	.027533 (.027975)	.030595 (.034070)	.032286 (.032716)	.027521 (.029065)
			M-Huber	.020573 (.020900)	<b>.023393</b> (.023445)	<b>.030462</b> (.032617)	<b>.024811</b> (.025775)	.021490 (.022365)
			M-Tukey	.068458 (.086371)	.051695 (.062963)	.064128 (.080539)	.068063 (.082042)	.067349 (.086128)
			M-Hampel	.044008 (.051485)	.037319 (.041772)	.043455 (.051074)	.044481 (.051566)	.043848 (.053287)
80	.05	5	Mean	<b>.013694</b> (.015032)	.027006 (.034676)	.091634 (.086135)	.020354 (.020679)	<b>.014521</b> (.014089)
			ITM	.027198 (.030876)	.024489 (.026528)	.027165 (.028146)	.025776 (.030491)	.026798 (.028607)
			DTM	.016380 (.017661)	.018634 (.021748)	.022894 (.025369)	.020462 (.021919)	.018528 (.018069)
			M-Huber	.014654 (.016522)	<b>.015973</b> (.018272)	<b>.022412</b> (.025206)	<b>.015169</b> (.015882)	.014894 (.014637)
			M-Tukey	.045315 (.057328)	.036955 (.044038)	.037720 (.042080)	.040191 (.049716)	.041766 (.051849)
			M-Hampel	.028004 (.032161)	.025747 (.028835)	.027647 (.028420)	.026892 (.032161)	.027624 (.031103)
50	.1	5	Mean	.068373 (.087502)	.338033 (.294301)	.050658 (.060348)	<b>.024532</b> (.023213)	
			ITM	.034176 (.038092)	<b>.037455</b> (.044174)	.038626 (.045606)	.042659 (.047163)	
			DTM	.032170 (.037113)	.053886 (.086974)	.042721 (.047632)	.033011 (.032304)	
			M-Huber	<b>.029538</b> (.031726)	.061060 (.073438)	<b>.029372</b> (.033033)	.025551 (.026263)	
			M-Tukey	.048841 (.054898)	.047792 (.054820)	.057261 (.074057)	.068544 (.086066)	
			M-Hampel	.036190 (.041169)	.037691 (.041771)	.041045 (.051050)	.044773 (.050343)	
80	.1	5	Mean	.040832 (.056402)	.297179 (.205967)	.027246 (.028767)	.014561 (.014493)	
			ITM	.022755 (.024800)	.021334 (.025422)	.021744 (.021647)	.025079 (.029365)	
			DTM	.018862 (.020951)	.034448 (.045930)	.024518 (.024682)	.018243 (.018276)	
			M-Huber	<b>.017224</b> (.019108)	.040752 (.044377)	<b>.016359</b> (.016532)	<b>.014327</b> (.015648)	
			M-Tukey	.030731 (.036051)	.028976 (.037778)	.032345 (.039253)	.041605 (.051620)	
			M-Hampel	.022850 (.024835)	<b>.021268</b> (.024713)	.022727 (.022942)	.026146 (.031736)	
50	.05	25	Mean	.608521 (.964104)	2.09387 (2.28607)	.362877 (.461968)	.060141 (.037809)	
			ITM	.039254 (.044935)	.035937 (.038311)	.042082 (.049751)	.037638 (.041441)	
			DTM	.029887 (.041196)	.028975 (.029200)	.125013 (.169009)	.077631 (.052955)	
			M-Huber	<b>.026531</b> (.032211)	<b>.028756</b> (.031740)	<b>.02526</b> (.027586)	<b>.023510</b> (.022924)	
			M-Tukey	.060952 (.078603)	.058171 (.070880)	.062487 (.074965)	.061798 (.079836)	
			M-Hampel	.042382 (.049789)	.040150 (.046511)	.043926 (.052464)	.040382 (.045112)	
80	.05	25	Mean	.417123 (.647261)	1.91755 (1.70766)	.223157 (.308306)	.037808 (.022049)	
			ITM	.026210 (.033108)	.024602 (.027218)	.023657 (.027654)	.02746 (.032988)	
			DTM	.017842 (.019961)	<b>.019117</b> (.018921)	.080250 (.094733)	.047756 (.030056)	
			M-Huber	<b>.016246</b> (.018939)	.020108 (.021302)	<b>.015948</b> (.016690)	<b>.015996</b> (.017307)	
			M-Tukey	.042438 (.056368)	.036947 (.051762)	.039888 (.052552)	.040606 (.049741)	
			M-Hampel	.026844 (.033460)	.025351 (.029268)	.024970 (.029514)	.027544 (.032193)	
50	.1	25	Mean	1.33172 (1.91994)	7.34928 (5.80978)	.649172 (.877986)	.100615 (.052521)	
			ITM	.035903 (.047082)	.036878 (.046616)	.035767 (.038587)	.032089 (.033875)	
			DTM	.044589 (.126620)	.238852 (.820844)	.281549 (.363469)	.122590 (.069193)	
			M-Huber	<b>.028959</b> (.030816)	.052190 (.058452)	<b>.027647</b> (.027143)	<b>.024738</b> (.024483)	
			M-Tukey	.050371 (.063786)	.046707 (.059544)	.049991 (.062703)	.046620 (.059655)	
			M-Hampel	.034459 (.039967)	<b>.034995</b> (.039768)	.037249 (.045652)	.034127 (.038973)	
80	.1	25	Mean	.850504 (1.27149)	6.83077 (4.31747)	.402271 (.468268)	.061555 (.027645)	
			ITM	.021967 (.024331)	.022761 (.033777)	.021311 (.024577)	.023471 (.026155)	
			DTM	.022268 (.038530)	.124036 (.550383)	.170359 (.218541)	.078601 (.038582)	
			M-Huber	<b>.019365</b> (.021481)	.039244 (.040842)	<b>.016147</b> (.018109)	<b>.016321</b> (.015291)	
			M-Tukey	.029648 (.034349)	.031106 (.040488)	.030552 (.039785)	.031939 (.039397)	
			M-Hampel	.022767 (.026150)	<b>.022316</b> (.026552)	.021796 (.025544)	.023852 (.027021)	

To evaluate the performance of the functional location estimators, we have calculated the integrated squared error for each of the  $N = 500$  samples. In models 1-5, the

integrated squared error (ISE) of a sample  $(x_1, \dots, x_n)$  is given by

$$ISE((x_1, \dots, x_n), \hat{g}) = \frac{1}{T} \sum_{k=1}^T \left[ \hat{g} \left( \frac{k}{T} \right) - g \left( \frac{k}{T} \right) \right]^2,$$

where  $\hat{g} = \hat{g}(x_1, \dots, x_n)$  can be any of the considered estimators (sample mean, functional location M-estimators using either Huber, Tukey or Hampel loss function, impartial trimmed mean or sample depth trimmed mean). Similarly, in models 6-9 the ISE of a sample becomes

$$ISE((x_1, \dots, x_n), \hat{g}) = \frac{1}{T} \sum_{k=1}^T \left[ \left( \hat{g} \left( \frac{k}{T} \right) - g \left( \frac{k}{T} \right) \right)^2 + \left( D\hat{g} \left( \frac{k}{T} \right) - Dg \left( \frac{k}{T} \right) \right)^2 \right].$$

To obtain the derivatives of the curves, we have used an approximation based on finite differences, but, naturally, other methods could be considered. Finally, to summarize the performance of the estimators, their mean squared error (MSE) together with its standard deviation have been computed as:

$$MSE(\hat{g}) = \frac{1}{N} \sum_{j=1}^N ISE(\hat{g}_j), \quad s = \sqrt{\frac{1}{N} \sum_{j=1}^N (ISE(\hat{g}_j) - MSE(\hat{g}))^2},$$

where  $ISE(\hat{g}_j)$  denotes the integrated squared error for estimator  $\hat{g}_j$  in sample  $j$  (where  $j = 1, \dots, N$ ).

The results of the simulations are collected in Table 1 for models 1-5 and in Table 2 for models 6-9. Similar results are obtained when another robust location estimator, the functional median based on the functional depth (Fraiman, R. and Muniz, G., 2001), is used to fix the tuning parameters of the loss functions and also to initialize the algorithm to compute the corresponding M-estimates.

Note that the column for model 1 in Table 1 only contains one set of results because this is the model without outliers. For each simulation setting, the lowest MSE among the different methods is shown in bold. From the results in these tables we can see that, as for the case of real- or vector-valued data, there is no uniformly best location estimator. For the non-contaminated case (model 1), the mean is naturally the best choice. However, the mean quickly deteriorates when the data contain contamination. For the considered settings, it turns out that the Huber M-estimator yields the lowest MSE in many situations, especially in the case of shape outliers (models 6-9 in Table 2). The Hampel M-estimator regularly yields the best results in model 3. Note that, of course, other choices of the tuning parameters for the Tukey and Hampel loss functions might lead to better results for these estimators. In any case, the current results already show the advantage of using M-estimators. These results confirm that M-estimators are robust functional location estimators which offer a good compromise between robustness (low bias) and efficiency, resulting in a low MSE generally.

**Table 2.** MSE and corresponding standard error (between brackets) of the mean, the trimmed mean (ITM), the depth trimmed mean (DTM), and Huber (M-Huber), Tukey (M-Tukey) and Hampel (M-Hampel) M-estimates in models 6-9

$n$	$\epsilon$	Estimator	Model 6	Model 7	Model 8	Model 9
50	.05	Mean	.278113 (.168456)	.321940 (.215465)	.275798 (.171702)	.319261 (.201474)
		ITM	.106459 (.096693)	.101038 (.087106)	.094702 (.083255)	.107983 (.095597)
		DTM	.390294 (.252391)	.461141 (.317118)	.387891 (.255572)	.446280 (.291377)
		M-Huber	<b>.069833</b> (.058318)	<b>.066822</b> (.055995)	<b>.064437</b> (.054801)	<b>.07176</b> (.059385)
		M-Tukey	.179066 (.169368)	.164969 (.144269)	.159793 (.156639)	.184133 (.183535)
		M-Hampel	.109636 (.104157)	.102254 (.088511)	.095988 (.087562)	.114649 (.102563)
80	.05	Mean	.171281 (.084480)	.207449 (.111521)	.163277 (.086557)	.201092 (.104882)
		ITM	.063997 (.057002)	.069350 (.061218)	.071373 (.071925)	.063025 (.059847)
		DTM	.240762 (.126659)	.295736 (.161671)	.229626 (.123639)	.292698 (.161348)
		M-Huber	<b>.042394</b> (.035933)	<b>.045964</b> (.039066)	<b>.045579</b> (.042751)	<b>.042162</b> (.036992)
		M-Tukey	.103399 (.098657)	.110678 (.106923)	.115176 (.111074)	.100845 (.096861)
		M-Hampel	.064670 (.059019)	.070369 (.061955)	.072703 (.073892)	.064386 (.061827)
50	.1	Mean	.503312 (.273895)	.569512 (.284861)	.493756 (.246616)	.584617 (.292074)
		ITM	.096455 (.087461)	.101523 (.084384)	.100245 (.087318)	.099889 (.086490)
		DTM	.699371 (.391043)	.814804 (.410607)	.703274 (.361495)	.858913 (.446574)
		M-Huber	<b>.077265</b> (.061095)	<b>.080585</b> (.062364)	<b>.074852</b> (.056917)	<b>.078349</b> (.062356)
		M-Tukey	.142503 (.139147)	.148784 (.135452)	.143118 (.143674)	.142575 (.140302)
		M-Hampel	.098219 (.088274)	.103504 (.085956)	.097889 (.087873)	.100827 (.085214)
80	.1	Mean	.316009 (.140514)	.366224 (.159513)	.307506 (.138383)	.372142 (.165260)
		ITM	.067109 (.058206)	.058388 (.048746)	.057200 (.049419)	.061645 (.053649)
		DTM	.450599 (.202575)	.530599 (.234825)	.431662 (.199880)	.541167 (.252168)
		M-Huber	<b>.052811</b> (.040578)	<b>.045542</b> (.034687)	<b>.046112</b> (.036250)	<b>.048030</b> (.037221)
		M-Tukey	.093819 (.092566)	.082335 (.078285)	.082007 (.079620)	.084871 (.082654)
		M-Hampel	.067228 (.058391)	.056552 (.047065)	.056812 (.051564)	.060219 (.050724)

## 5. Real-data example

We illustrate the M-estimates of location for functional data with the following example.

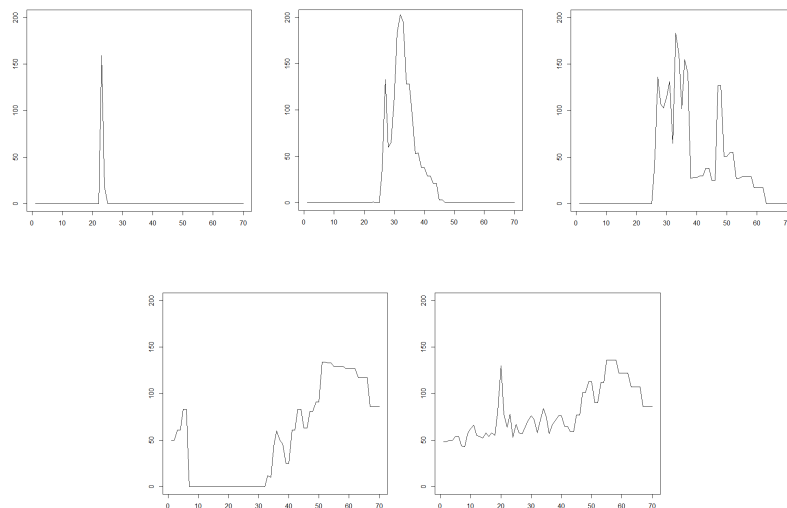
Consider the dataset of  $n = 472$  radar waves registered by the satellite Topex/Poseidon around an area of 25 kilometers upon the Amazon River, with the aim of using them for altimetric and hydrological purposes. The space of study is  $\mathbb{H} = L^2([0, 70])$ . The dataset contains observed values of these curves at 70 equidistant time points within the range  $[0, 70]$ . The data set, together with a brief description, can be obtained from the web page <http://www.math.univ-toulouse.fr/staph/npfda/npfda-datasets>. More detailed information about the data can be found in [Frappart, F. \(2003\)](#).

As outlined on the web page of the data and shown in Figure 2, there are different types of waves. Figure 2 displays the linearly interpolated measurements for some of the waves, to show the differences in the types of waves in this data set. Namely, there are

- curves with one heavy peak, like curve number 21;
- curves with one less heavy peak, like curve number 3;
- curves that seem to have more than one peak, like curve number 1;
- curves that do not seem to have a real peak, like curve number 5;
- ‘flat noisy curves’, like curve number 4.

The data set is thus heterogeneous and thus we use robust M-estimators to estimate the center of the majority of the functional data. For this analysis, we use the  $L^2$  norm, so we focus on magnitude outliers. We consider the Huber, Tukey and Hampel loss functions. The tuning constants in these loss functions are determined as explained before, using





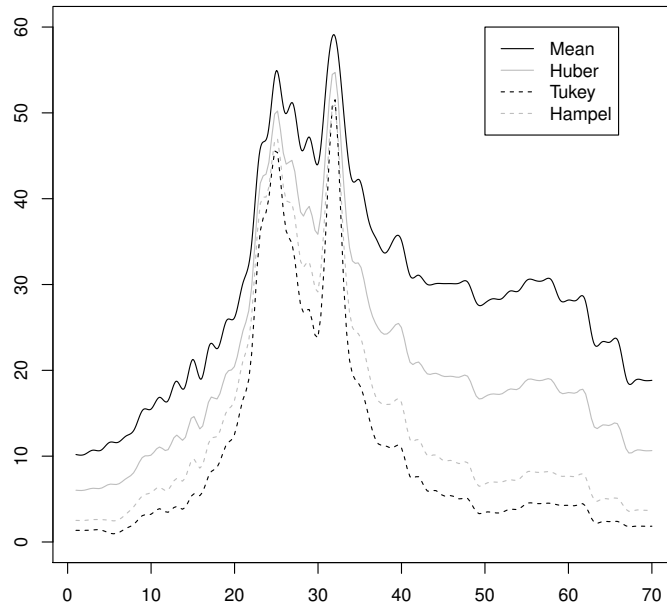
**Figure 2.** Waveforms numbers 21 (top left), 3 (top middle), 1 (top right), 5 (bottom left) and 4 (bottom right).

the impartial trimmed mean (with  $\alpha = 0.2$ ) as initial solution.

The three resulting M-estimates are plotted in Figure 3, where the functional sample mean has been added as well. Note that the algorithm for the functional M-estimators as explained in Section 2 obtains the values of these location M-estimates evaluated on the 70 equidistant points. The algorithm does not require any pre-smoothing of the functional observations and thus the resulting estimates cannot be influenced by such preprocessing of the data. However, to represent the resulting M-estimates as curves, some post-smoothing of the estimates returned by the algorithm is needed. For this purpose we have used the function `Data2fd` from the R package `fda`, considering a B-spline basis in this example. The obtained estimated curves are shown in Figure 3. From this figure we can clearly see that the sample mean is highly influenced by the deviating types of waves (i.e., flat noisy curves), whereas Huber and especially Tukey and Hampel M-estimates show a more robust behavior, leading to a better estimate of the center of the majority of the curves.

## 6. Concluding remarks

We have introduced M-estimators of functional location and studied their properties. Well-known loss functions such as Huber, Tukey and Hampel loss functions yield robust location estimators also in the functional context. The Representer Theorem allows us to represent these M-estimators as adaptively weighted means. Since the estimators are



**Figure 3.** Estimates for the center of the radar waves data, obtained by the mean and Huber, Tukey and Hampel M-estimators.

not scale equivariant in general, the tuning constant(s) in these loss functions need to be selected with some care. We have used the distribution of distances corresponding to the impartial trimmed mean for functional data for this purpose. The impartial trimmed mean has also been used as initial estimate in the iterative algorithm that is used to calculate the M-estimates. The simulation study suggests that the Huber M-estimator of functional location shows good behavior in a wide range of contamination settings. However, the example indicates that the Tukey and Hampel M-estimators, which use a bounded loss function, may yield more robust results in some (more extreme) contamination scenarios.

In future work, we will focus on extending these results to important related settings such as multivariate functional data and (fuzzy) set-valued data which can be embedded in convex cones of certain Hilbert spaces (see e.g. [González-Rodríguez, G. et al., 2012](#); [Sinova, B. et al., 2014](#)). Moreover, alternative sets of sufficient conditions guaranteeing the strong consistency and Borel measurability of M-estimators of location could be searched for in order to relax the requirements for strong consistency to include infinite

dimensional spaces. Finally, a thorough study on the selection of the tuning parameters and their influence on the estimates should be developed.

## Acknowledgements

The authors are grateful to the reviewers and Editor handling this paper for their insightful comments and suggestions. This research has been partially supported by the Spanish Ministry of Economy and Competitiveness Grant MTM2013-44212-P, the Principality of Asturias/FEDER Grants GRUPIN14-101, Grant C16/15/068 of International Funds KU Leuven and IAP research network grant nr. P7/06 of the Belgian government.

## References

- Beaton, A.E. and Tukey, J.W. (1974) The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, **16**, 147–185.
- Brown, L.D. and Purves, R. (1973) Measurable selections of extrema. *Ann. Statist.*, **1**(5), 902–912.
- Chichignoud, M. and Lederer, J. (2014) A robust, adaptive M-estimator for pointwise estimation in heteroscedastic regression. *Bernoulli*, **20**(3), 1560–1599.
- Christmann, A. and Steinwart, I. (2007) Consistency and robustness of kernel-based regression in convex risk minimization. *Bernoulli*, **13**(3), 799–819.
- Ciarleglio, A. and Ogden, R.T. (2016) Wavelet-based scalar-on-function finite mixture regression models. *Comp. Stat. Data Anal.*, **93**, 86–96.
- Claeskens, G., Hubert, M., Slaets, L. and Vakili, K. (2014) Multivariate functional halfspace depth. *J. Amer. Stat. Assoc.*, **109**, 411–423.
- Colubi, A. and González-Rodríguez, G. (2015) Fuzziness in data analysis: Towards accuracy and robustness. *Fuzzy Sets Syst.*, **281**, 260–271.
- Cuesta-Albertos, J.A. and Fraiman, R. (2007) Impartial trimmed means for functional data. *Comp. Stat. Data Anal.*, **51**(10), 4864–4877.
- Cuesta-Albertos, J.A. and Nieto-Reyes, A. (2008) The random Tukey depth. *Comp. Stat. Data Anal.*, **52**(11), 4979–4988.
- Cuevas, A., Febrero, M. and Fraiman, R. (2007) Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, **22**, 481–496.
- Cuevas, A. and Fraiman, R. (2009) On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivar. Anal.*, **100**, 753–766.
- Domingues, M.A.O., De Souza, R.M.C.R. and Cysneiros, F.J.A. (2010) A robust method for linear regression of symbolic interval data. *Patt. Recog. Let.*, **31**(13), 1991–1996.
- Donoho, D.L. and Huber, P.J. (1983) The notion of breakdown point. In: *A Festschrift for Erich L. Lehmann* (Bickel, P.J., Doksum, K. and Hodges, J.L. Jr, Eds.), pp. 157–184. Belmont: Wadsworth.
- D’Urso, P. and De Giovanni, L. (2014) Robust clustering of imprecise data. *Chem. Intel. Lab. Syst.*, **136**, 58–80.

- D'Urso, P., De Giovanni, L. and Massari, R. (2015) Trimmed fuzzy clustering for interval-valued data. *Adv. Data Anal. Class.*, **9**, 21–40.
- D'Urso, P., Massari, R. and Santoro, A. (2011) Robust fuzzy regression analysis. *Inform. Sci.*, **181**(19), 4154–4174.
- Ferraty, F. and Vieu, P. (2006) *Non Parametric Functional Data Analysis. Theory and Practice*. Berlin: Springer.
- Filzmoser, P. and Todorov, V. (2013) Robust tools for the imperfect world. *Inform. Sci.*, **245**, 4–20.
- Fraiman, R. and Muniz, G. (2001) Trimmed means for functional data. *Test*, **10**(2), 419–440.
- Fraiman, R. and Svarc, M. (2013) Resistant estimates for high dimensional and functional data based on random projections. *Comp. Stat. Data Anal.*, **58**, 326–338.
- Frappart, F. (2003) *Catalogue des formes d'onde de l'altimètre Topex/Poséidon sur le bassin amazonien*. Tec. Rep., CNES, Toulouse, France.
- Fritz, H., García-Escudero, L.A. and Mayo-Iscar, A. (2013) Robust constrained fuzzy clustering. *Inform. Sci.*, **245**, 38–52.
- González, J., Vujačić, I. and Wit, E. (2014) Reproducing kernel Hilbert space based estimation of systems of ordinary differential equations. *Patt. Recogn. Let.*, **45**(1), 26–32.
- González-Rodríguez, G., Colubi, A. and Gil, M.A. (2012) Fuzzy data treated as functional data. A one-way ANOVA test approach. *Comp. Stat. Data Anal.*, **56**, 943–955.
- Hampel, F.R. (1971) A general qualitative definition of robustness. *Ann. Math. Stat.*, **42**(6), 1887–1896.
- Hampel, F.R. (1974) The influence curve and its role in robust estimation. *J. Amer. Stat. Assoc.*, **69**, 383–393.
- Hampel, F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986) *Robust Statistics: the Approach Based on Influence Functions*. New York: John Wiley and Sons.
- Hsing, T. and Eubank, R. (2015) *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Chichester: J. Wiley & Sons.
- Hu, Q., An, S., Yu, X. and Yu, D. (2011) Robust fuzzy rough classifiers. *Fuzzy Sets Syst.*, **183**(1), 26–43.
- Huber, P.J. (1964) Robust estimation of a location parameter. *Ann. Math. Statist.*, **35**(1), 73–101.
- Huber, P.J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proc. 5th Berkeley Symp. Math. Stat. & Prob.*, **1**, 221–233.
- Huber, P.J. (1981) *Robust Statistics*. New York: J. Wiley & Sons.
- Hubert, M., Rousseeuw, P. and Segaert, P. (2015) Multivariate functional outlier detection. *Stat. Methods Appl.*, **24**, 177–202.
- Jacques, J. and Preda, C. (2014) Model-based clustering for multivariate functional data. *Comp. Stat. Data Anal.*, **71**, 92–106.
- Jerrard, R.L. and Sternberg, P. (2009) Critical points via  $\Gamma$ -convergence: general theory and applications. *J. Eur. Math. Soc.*, **11**, 705–753.
- Kim, J.S. (2011). *Kernel Methods for Classification with Irregularly Sampled and Contaminated Data*. PhD Thesis, University of Michigan ([http://deepblue.lib.umich.edu/bitstream/handle/2027.42/89858/stannum\\_1.pdf?sequence=1](http://deepblue.lib.umich.edu/bitstream/handle/2027.42/89858/stannum_1.pdf?sequence=1))
- Kim, J.S. and Scott, C.D. (2011) On the robustness of kernel density M-estimators. In: *Proc.*

- 28th Int. Conf. Mach. Lear. Bellevue, Washington, 697–704.
- Kim, J.S. and Scott, C.D. (2012) Robust kernel density estimation. *J. Mach. Learn. Res.*, **13**, 2529–2565.
- López-Pintado, S. and Romo, J. (2009) On the concept of depth for functional data. *J. Amer. Stat. Assoc.*, **104** (486), 718–734.
- López-Pintado, S. and Romo, J. (2011) A half-region depth for functional data. *Comp. Stat. Data Anal.*, **55**, 1679–1695.
- Lopuhaä, H.P. and Rousseeuw, P.J. (1991) Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Stat.*, **19**, 229–248.
- Ma, J., Zhao, J., Tian, J., Yuille, A.L. and Tu, Z. (2014) Robust point matching via vector field consensus. *IEEE Trans. Im. Proc.*, **23**(4), 1706–1721.
- Maronna, R.A., Martin, D.R. and Yohai, V.J. (2006) *Robust Statistics: Theory and Methods*. New York: J. Wiley & Sons.
- Minsker, S. (2015) Geometric median and robust estimation in Banach spaces. *Bernoulli*, **21**(4), 2308–2335.
- Ordóñez-Cabrera, M., Rosalsky, A. and Volodin, A. (2012) Some theorems on conditional mean convergence and conditional almost sure convergence for randomly weighted sums of dependent random variables. *Test*, **21**(2), 369–385.
- Paulauskas, V. and Račkauskas, A. (1989) *Approximation theory in the central limit theorem*. Dordrecht: Springer.
- Ramsay, J. and Silverman, B.W. (2005). *Functional Data Analysis* (2nd edn.). New York: Springer.
- Rockafellar, R.T. and Wets, R.J.B. (1998). *Variational Analysis*. Heidelberg: Springer.
- Rosen, B.R., Buckner, R.L. and Dale, A.M. (1998) Event-related functional MRI: Past, present, and future. *Proc. Natl. Acad. Sci. U.S.A.*, **95**(3), 773–780.
- Rousseeuw, P.J. and Van Driessen, K. (2006) Computing LTS regression for large data sets. *Data Min. Know. Disc.*, **12**(1), 29–45.
- Sinova, B., Gil, M.A., Colubi, A. and Van Aelst, S. (2012) The median of a random fuzzy number. The 1-norm distance approach. *Fuzzy Sets Syst.*, **200**, 99–115.
- Sinova, B., Gil, M.A., López, M.T. and Van Aelst, S. (2014) A parameterized  $L^2$  metric between fuzzy numbers and its parameter interpretation. *Fuzzy Sets Syst.*, **245**, 101–115.
- Sinova, B., Pérez-Fernández and S., Montenegro, M. (2015) The Wabl/Ldev/Rdev Median of a Random Fuzzy Number and Statistical Properties. In: *Strengthening Links Between Data Analysis and Soft Computing* (Grzegorzewski, P., Gagolewski, M., Hryniewicz, O. and Gil, M.A., Eds.), pp. 143–150. Heidelberg: Springer.
- Sreenivasa, M., Souères, P. and Laumond, J.P. (2012) Walking to grasp: modeling of human movements as invariants and an application to humanoid robotics. *IEEE Trans. Syst., Man, Cybern. A*, **42**(4) 880–893.
- Steinwart, I. and Christmann, A. (2008) *Support Vector Machines*. New York: Springer.
- Szilágyi, L. (2013) Robust spherical shell clustering using fuzzy-possibilistic product partition. *Int. J. Intel. Syst.*, **28**(6), 524–539.
- Vakhania, N., Tarieladze, V. and Chobanyan, S. (1987) *Probability distributions on Banach spaces*. Dordrecht: Springer Science & Business Media.
- Vandermeulen, R.A. and Scott, C.D. (2013) Consistency of robust kernel density estimators.

- JMLR: Work. Conf. Proc.*, **30**, 1–24.
- Winkler, R., Klawonn, F. and Kruse, R. (2011) Fuzzy clustering with polynomial fuzzifier function in connection with M-estimators. *Appl. Comput. Math.*, **10**(1) 2011, 146–163.
- Wood, A.T.A. and Chan, G. (1994) Simulation of stationary Gaussian processes in  $C[0, 1]$ . *J. Comput. Graphics Stat.*, **3**(4), 409–432.
- Yohai, V.J. (1987) High breakdown-point and high efficiency robust estimates for regression. *Ann. Statist.*, **15**, 642–656.
- Zhang, Z. and Zhang, Y. (2014) Variable kernel density estimation based robust regression and its applications. *Neurocomputing*, **134**, 30–37.