

# Scale equivariant alternative for fuzzy M-estimators of location

Beatriz Sinova<sup>1</sup>

**Abstract** The Aumann-type mean fulfills very convenient properties as a location measure of a random fuzzy number, but its high sensitivity to outliers makes other alternatives, such as fuzzy M-estimators of location, more suitable to describe contaminated data sets. Under some conditions, fuzzy M-estimators fulfill properties such as the strong consistency and the translation equivariance. However, the scale equivariance does not hold in general and the choice of the measurement units may have too much influence on the results. A first solution to solve this was the selection of the tuning parameters involved in the most used loss functions (Huber's, Tukey's and Hampel's) in terms of the distribution of distances of the observed data to the considered initial location estimate. Now a second solution is proposed including a robust estimate of the unknown dispersion in the definition of fuzzy M-estimators of location. The empirical comparison of both proposals shows that the latter solution may be more suitable for dealing with extreme data, and therefore it could better identify which observations should be considered outliers indeed.

---

Departamento de Estadística e I.O. y D.M., Universidad de Oviedo, Spain  
sinovabeatriz@uniovi.es

This paper is dedicated to the memory of Prof. Pedro Gil, who not only taught my mother and I Statistics in an interesting and calm way, but also left me bright memories in relation to our condition of neighbours and the conferences we both attended, as well as the *Champanadas* he cheered up with his accordion. I am deeply grateful for such moments and lessons.

## 1 Introduction

Fuzzy numbers can model experiments characterized by an underlying imprecision, such as ratings, opinions or perceptions (see e.g. De la Rosa de Sáa *et al.* [3]). Due to their interest, statistical methodology is being adapted to analyze this kind of data. With respect to central tendency measures, the best-known one is the Aumann-type mean [8], which generalizes the notion of mean of a real-valued random variable. Even when the Aumann-type mean fulfills very good statistical and probabilistic properties, outliers have too much impact on its estimate. For that reason, M-estimators of location have been recently defined in the fuzzy number-valued case by Sinova *et al.* [10] and their robustness has been shown. This is not the only robust location measure for fuzzy numbers proposed in the literature (we could think, for example, about some extensions of the concept of median like the ones introduced in [11] and [12]), but their performance seems to be the best in general. Although the empirical study addressed in [10] concludes that there is no uniformly best location estimator, it highlights the good behavior of fuzzy M-estimators of location.

Scale equivariance, on the contrary, does not hold for fuzzy M-estimators of location unless the loss function involved in their definition is a power function. This is an important drawback, since it means that measurement units could have a lot of impact on the results. A first solution was provided in [10] and consists in choosing the tuning parameters in the used loss functions (Huber's and Hampel's, and it could be also applied to other functions such as Tukey's) taking into account the distribution of distances from the observed data to the initial estimate considered for the computation of the corresponding M-estimator. Therefore, the loss function is adapted to the magnitude of the data we are working with.

The aim of this paper is to present an alternative to solve the lack of scale equivariance. In the classical settings, where the same problem has had to be dealt with, a robust estimate of the dispersion is introduced in the definition of the M-estimator of location to make it scale equivariant. Recently, a robust estimate of the dispersion of a random fuzzy number, the median distance deviation about the median, has been analyzed (see [4]). The idea is, in consequence, to use a similar median distance deviation about the median to extend M-estimators of location with unknown dispersion to the fuzzy number-valued settings.

The rest of the paper is structured as follows. The preliminaries on the space of fuzzy numbers and fuzzy M-estimators of location are recalled in Section 2. Section 3 presents the concept of fuzzy M-estimators of location with unknown dispersion and the study of their scale equivariance, whereas their empirical comparison with the previously defined fuzzy M-estimators of location is presented in Section 4. Finally, some concluding remarks are provided in Section 5.

## 2 Preliminaries on the space of fuzzy numbers and fuzzy M-estimation of location

In this section, the most important characteristics of the space of fuzzy numbers will be recalled, as well as the adaptation of M-estimators of location to the fuzzy-valued settings.

$\mathcal{F}_c(\mathbb{R})$  will denote the class of (bounded) fuzzy numbers, which are mappings  $\tilde{U} : \mathbb{R} \rightarrow [0, 1]$  such that their  $\alpha$ -levels

$$\tilde{U}_\alpha = \begin{cases} \{x \in \mathbb{R} : \tilde{U}(x) \geq \alpha\} & \text{if } \alpha \in (0, 1] \\ \text{cl}\{x \in \mathbb{R} : \tilde{U}(x) > 0\} & \text{if } \alpha = 0, \end{cases}$$

are nonempty compact intervals. It is possible to interpret  $\tilde{U}(x)$  as the ‘degree of compatibility’ of  $x$  with  $\tilde{U}$  (or ‘degree of truth’ of the assertion “ $x$  is  $\tilde{U}$ ”).

Fuzzy data are very useful to model those phenomena such as human perceptions or valuations which present an underlying imprecision. Indeed, their  $\alpha$ -levels incorporate a certain gradualness that does not appear when dealing with interval-valued data.

Concerning the mathematical operations among these kinds of data, the sum and the product are defined by means of Zadeh’s extension principle, which extends level-wise the usual interval arithmetic.

**Definition 1.** Let  $\tilde{U}, \tilde{V} \in \mathcal{F}_c(\mathbb{R})$ . The **sum** of  $\tilde{U}$  and  $\tilde{V}$  is defined as the fuzzy number  $\tilde{U} + \tilde{V} \in \mathcal{F}_c(\mathbb{R})$  given for each  $\alpha \in [0, 1]$  by

$$(\tilde{U} + \tilde{V})_\alpha = \text{Minkowski sum of } \tilde{U}_\alpha \text{ and } \tilde{V}_\alpha = [\inf \tilde{U}_\alpha + \inf \tilde{V}_\alpha, \sup \tilde{U}_\alpha + \sup \tilde{V}_\alpha].$$

Let  $\tilde{U} \in \mathcal{F}_c(\mathbb{R})$  and  $\gamma \in \mathbb{R}$ . The **product** of  $\tilde{U}$  by the scalar  $\gamma$  is defined as the fuzzy number  $\gamma \cdot \tilde{U} \in \mathcal{F}_c(\mathbb{R})$  given for each  $\alpha \in [0, 1]$  by

$$(\gamma \cdot \tilde{U})_\alpha = \gamma \cdot \tilde{U}_\alpha = \begin{cases} [\gamma \cdot \inf \tilde{U}_\alpha, \gamma \cdot \sup \tilde{U}_\alpha] & \text{if } \gamma \geq 0, \\ [\gamma \cdot \sup \tilde{U}_\alpha, \gamma \cdot \inf \tilde{U}_\alpha] & \text{otherwise.} \end{cases}$$

Now the family of distances between fuzzy numbers introduced by Montenegro *et al.* [7], which extends the one proposed by Bertoluzza *et al.* [1], will be recalled. Note that its use is very convenient due to the lack of linearity in the space  $(\mathcal{F}_c(\mathbb{R}), +, \cdot)$  as explained in e.g. [2].

**Definition 2.** Let  $\theta \in (0, +\infty)$  and let  $\varphi$  be an absolutely continuous probability measure on  $([0, 1], \mathcal{B}_{[0,1]})$  with the mass function being positive on  $(0, 1)$ . The **mid/spr-based  $L^2$  distance** between any two fuzzy numbers  $\tilde{U}, \tilde{V} \in \mathcal{F}_c(\mathbb{R})$  is defined as

$$D_\theta^\varphi(\tilde{U}, \tilde{V}) = \left[ \int_{[0,1]} \left( \text{mid } \tilde{U}_\alpha - \text{mid } \tilde{V}_\alpha \right)^2 + \theta \left( \text{spr } \tilde{U}_\alpha - \text{spr } \tilde{V}_\alpha \right)^2 d\varphi(\alpha) \right]^{1/2},$$

where  $\text{mid } \tilde{U}_\alpha = (\inf \tilde{U}_\alpha + \sup \tilde{U}_\alpha)/2$  and  $\text{spr } \tilde{U}_\alpha = (\sup \tilde{U}_\alpha - \inf \tilde{U}_\alpha)/2$ .

The role of  $\theta$  and  $\varphi$  is not stochastic, but to weigh the importance of the deviation ‘in shape’ in contrast to the deviation ‘in center’, and the relevance of the different  $\alpha$ -levels, respectively. It can be proven that the usual choice  $1/3$  makes all the points in the intervals (once fixed any  $\alpha$ ) equally important.

M-estimators of location will be defined in terms of the mid/spr-based  $L^2$  distance since the space  $(\mathcal{F}_c(\mathbb{R}), D_\theta^\varphi)$  can be isometrically embedded into a convex cone of a certain Hilbert space by means of the so-called support function (in Puri and Ralescu’s sense [8]).

The notion of random fuzzy number in Puri and Ralescu’s sense [8] mathematically formalizes the random mechanism generating fuzzy data.

**Definition 3.** Let  $(\Omega, \mathcal{A}, P)$  be a probability space modeling a random experiment. A mapping  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$  is said to be a **random fuzzy number** associated with the random experiment if, and only if, for each  $\alpha \in [0, 1]$  the interval-valued mapping  $\mathcal{X}_\alpha$  (where  $\mathcal{X}_\alpha(\omega) = (\mathcal{X}(\omega))_\alpha$  for all  $\omega \in \Omega$ ) is a random compact interval or equivalently, the real-valued functions  $\inf \mathcal{X}_\alpha$  and  $\sup \mathcal{X}_\alpha$  are random variables.

A random fuzzy number is Borel-measurable with respect to the Borel  $\sigma$ -field associated with the  $D_\theta^\varphi$  distance. In order to summarize the central tendency of a random fuzzy number, one of the best-known measures is the following.

**Definition 4.** Let  $\mathcal{X}$  be a random fuzzy number and assume that the expected values of the random variables  $\inf \mathcal{X}_0$  and  $\sup \mathcal{X}_0$  are finite. The **Aumann-type mean** of  $\mathcal{X}$  is the fuzzy number  $\tilde{E}(\mathcal{X}) \in \mathcal{F}_c(\mathbb{R})$  such that for each  $\alpha \in [0, 1]$

$$(\tilde{E}(\mathcal{X}))_\alpha = [E(\inf \mathcal{X}_\alpha), E(\sup \mathcal{X}_\alpha)].$$

As an extension of the concept of mean for real-valued random variables, the Aumann-type mean inherits very convenient statistical and probabilistic properties, but also the high sensitivity to outliers. For this reason, other location measures for fuzzy-valued data with a more robust behavior have already been proposed in the literature. In particular, the concept of median has been extended to the fuzzy number-valued settings as follows.

**Definition 5.** Let  $\mathcal{X}$  be a random fuzzy number and  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ . The (sample) **1-norm median** is the fuzzy number  $\widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)$ , for short  $\widehat{\text{Me}}$ , such that for each  $\alpha \in [0, 1]$  it coincides with

$$[\text{Me}\{\inf(\mathcal{X}_1)_\alpha, \dots, \inf(\mathcal{X}_n)_\alpha\}, \text{Me}\{\sup(\mathcal{X}_1)_\alpha, \dots, \sup(\mathcal{X}_n)_\alpha\}],$$

with Me denoting the median of a real-valued random variable. In case any of the medians is non-unique, the convention of considering the midpoint of the interval of possible medians is used.

Among the robust location measures for fuzzy-valued data, the performance of fuzzy M-estimators of location is certainly remarkable, achieving the best results in many of the situations studied in [10].

**Definition 6.** Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$  be an associated random fuzzy number. Moreover, let  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ . Then, the **fuzzy M-estimator of location** is the fuzzy number-valued statistic  $\widehat{g}^M(\mathcal{X}_1, \dots, \mathcal{X}_n)$ , given by

$$\widehat{g}^M(\mathcal{X}_1, \dots, \mathcal{X}_n) = \arg \min_{\tilde{U} \in \mathcal{F}_c(\mathbb{R})} \frac{1}{n} \sum_{i=1}^n \rho(D_\theta^\varphi(\mathcal{X}_i, \tilde{U})),$$

if it exists, where the loss function  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}$  is assumed to be continuous and non-decreasing and to vanish at 0.

In [10, 9] it has been proven that, even when they may fulfill very good properties, fuzzy M-estimators of location are not scale equivariant unless  $\rho$  is a power function, which is not a possible choice if we are looking for robustness. [10] selects the tuning parameters involved in some well-known loss functions (Huber's and Hampel's) depending on the distribution of distances from the observed data to the initial estimate considered for the computation of the M-estimator to avoid the bad influence of the measurement units. A second alternative is introduced in Section 3 of this paper by extending the classical M-estimators of location with unknown dispersion, which are based on a robust estimate of the dispersion. The median distance deviation about the median of a random fuzzy number has been defined using the  $\rho_1$  distance, which is an  $L^1$  metric based on the infimum/supremum characterization of fuzzy numbers, in [4]. However, a new alternative is now considered, by replacing the  $\rho_1$  distance by the  $D_\theta^\varphi$  metric, since fuzzy M-estimators of location are defined in terms of the latter (due to the isometrical embedding mentioned above).

**Definition 7.** Let  $\mathcal{X}$  be a random fuzzy number and  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ . The (sample) **median  $D_\theta^\varphi$ -distance deviation about the 1-norm median** (MDD) is the following real number

$$\widehat{\sigma}_\theta^\varphi(\mathcal{X}_1, \dots, \mathcal{X}_n) = \text{Me} \left\{ D_\theta^\varphi(\mathcal{X}_1, \widehat{\text{Me}}), \dots, D_\theta^\varphi(\mathcal{X}_n, \widehat{\text{Me}}) \right\},$$

applying the same convention as in Definition 5.

It can be proven that the median  $D_\theta^\varphi$ -distance deviation about the 1-norm median satisfies the scale equivariance property.

**Proposition 1.** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$  be an associated random fuzzy number. Moreover, let  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ . The median  $D_\theta^\varphi$ -distance deviation about the 1-norm median is scale equivariant, that is, given any  $\gamma \in \mathbb{R}$ ,*

$$\widehat{\sigma}_\theta^\varphi(\gamma \cdot \mathcal{X}_1, \dots, \gamma \cdot \mathcal{X}_n) = |\gamma| \cdot \widehat{\sigma}_\theta^\varphi(\mathcal{X}_1, \dots, \mathcal{X}_n).$$

*Proof* First, due to the properties of the 1-norm median (see [12]),

$$\widehat{\text{Me}}(\gamma \cdot \mathcal{X}_1, \dots, \gamma \cdot \mathcal{X}_n) = \gamma \cdot \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n).$$

Therefore, for all  $i \in \{1, \dots, n\}$ ,

$$D_\theta^\varphi(\gamma \cdot \mathcal{X}_i, \widehat{\text{Me}}(\gamma \cdot \mathcal{X}_1, \dots, \gamma \cdot \mathcal{X}_n)) = |\gamma| \cdot D_\theta^\varphi(\mathcal{X}_i, \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n))$$

since  $\text{mid}(\gamma \cdot \widetilde{U})_\alpha = \gamma \cdot \text{mid} \widetilde{U}_\alpha$  and  $\text{spr}(\gamma \cdot \widetilde{U})_\alpha = |\gamma| \cdot \text{spr} \widetilde{U}_\alpha$  for all  $\widetilde{U} \in \mathcal{F}_c(\mathbb{R})$  and all  $\alpha \in [0, 1]$ . Finally,

$$\begin{aligned} & \widehat{\sigma}_\theta^\varphi(\gamma \cdot \mathcal{X}_1, \dots, \gamma \cdot \mathcal{X}_n) \\ &= \text{Me} \left\{ D_\theta^\varphi(\gamma \cdot \mathcal{X}_1, \widehat{\text{Me}}(\gamma \cdot (\mathcal{X}_1, \dots, \mathcal{X}_n))), \dots, D_\theta^\varphi(\gamma \cdot \mathcal{X}_n, \widehat{\text{Me}}(\gamma \cdot (\mathcal{X}_1, \dots, \mathcal{X}_n))) \right\} \\ &= \text{Me} \left\{ |\gamma| \cdot D_\theta^\varphi(\mathcal{X}_1, \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)), \dots, |\gamma| \cdot D_\theta^\varphi(\mathcal{X}_n, \widehat{\text{Me}}(\mathcal{X}_1, \dots, \mathcal{X}_n)) \right\} \\ &= |\gamma| \cdot \widehat{\sigma}_\theta^\varphi(\mathcal{X}_1, \dots, \mathcal{X}_n). \end{aligned}$$

□

### 3 Location M-estimators with unknown dispersion for random fuzzy numbers

Due to the lack of scale equivariance of fuzzy M-estimators of location, the classical M-estimators of location with unknown dispersion will be now extended to the fuzzy number-valued case. The alternative of simultaneously estimating both the location and scale will be not considered in this paper, since this procedure is already not satisfactory in the classical settings from the robustness point of view and also due to the numerical inconvenience of solving the system of two non-linear equations.

**Definition 8.** Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$  be an associated random fuzzy number. Moreover, let  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple

random sample from  $\mathcal{X}$ . Then, the **fuzzy MDD-based M-estimator of location** is the fuzzy number-valued statistic  $\widehat{g_{MDD}^M}(\mathcal{X}_1, \dots, \mathcal{X}_n)$ , given by

$$\widehat{g_{MDD}^M}(\mathcal{X}_1, \dots, \mathcal{X}_n) = \arg \min_{\tilde{U} \in \mathcal{F}_c(\mathbb{R})} \frac{1}{n} \sum_{i=1}^n \rho \left( \frac{D_{\tilde{\theta}}^{\varphi}(\mathcal{X}_i, \tilde{U})}{\widehat{\sigma}_{\tilde{\theta}}^{\varphi}(\mathcal{X}_1, \dots, \mathcal{X}_n)} \right),$$

if it exists, where the loss function  $\rho : \mathbb{R}^+ \rightarrow \mathbb{R}$  is assumed to be continuous and non-decreasing and to vanish at 0.

It can be shown that the fuzzy MDD-based M-estimator of location is indeed scale equivariant as it happens in the real-valued settings.

**Proposition 2.** *Let  $(\Omega, \mathcal{A}, P)$  be a probability space and  $\mathcal{X} : \Omega \rightarrow \mathcal{F}_c(\mathbb{R})$  be an associated random fuzzy number. Moreover, let  $(\mathcal{X}_1, \dots, \mathcal{X}_n)$  be a simple random sample from  $\mathcal{X}$ . The fuzzy MDD-based M-estimator of location is scale equivariant, that is, given any  $\gamma \in \mathbb{R}$ ,*

$$\widehat{g_{MDD}^M}(\gamma \cdot \mathcal{X}_1, \dots, \gamma \cdot \mathcal{X}_n) = |\gamma| \cdot \widehat{g_{MDD}^M}(\mathcal{X}_1, \dots, \mathcal{X}_n).$$

## 4 Simulation study

This section aims to empirically compare the two alternatives proposed to extend M-estimators of location to the fuzzy-valued case and avoid any problem with the measurement units. First, the tuning parameters involved in the most used loss functions, such as Huber's, Tukey's or Hampel's could be selected in terms of the distribution of distances of the observed data to the considered initial location estimate as in [10]. Secondly, fuzzy MDD-based M-estimators of location are a scale equivariant measure that has been introduced in Section 3.

Among the usual loss functions, we will consider the Hampel loss function since its suitability was shown in [10] for many of the studied cases. The *Hampel loss function* [5] corresponds to

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

where the nonnegative parameters  $a < b < c$  allow us to control the degree of suppression of large errors. The smaller their values, the higher this degree.

Note that apart from not being convex, this function can cope with extreme outliers, since observations far from the center ( $x \geq c$ ) all contribute equally to the loss.

Recall that the first alternative consists of fuzzy M-estimators of location carefully choosing the values of the tuning parameters. Following Kim and Scott [6], we will take  $a$ ,  $b$  and  $c$  as the median, 75th and 85th percentiles of the distances between the observations and an initial estimate, which throughout this paper will be the 1-norm median.

100 trapezoidal fuzzy data are generated according to four real-valued random variables:  $\mathcal{X} = \text{Tra}(X_1 - X_2 - X_3, X_1 - X_2, X_1 + X_2, X_1 + X_2 + X_4)$ , so  $\inf \mathcal{X}_0 = X_1 - X_2 - X_3$ ,  $\inf \mathcal{X}_1 = X_1 - X_2$ ,  $\sup \mathcal{X}_1 = X_1 + X_2$  and  $\sup \mathcal{X}_0 = X_1 + X_2 + X_4$ .

A contamination proportion equal to  $c_p \in \{0, 0.1, 0.2, 0.4\}$  is introduced in each sample. Any kind of outlier is allowed in these simulation studies: all the random variables detailed above ( $X_1, X_2, X_3$  and  $X_4$ ) can follow the corresponding distributions for the contaminated observations or just some (at least one) of them. This means that we deal with outliers in location, outliers in shape and/or outliers in both location and shape. A second parameter,  $C_D \in \{0, 1, 5, 10, 100\}$ , determines the distance between the distribution of the regular and contaminated observations.

In CASE 1 the variables  $X_i$  are independent. In particular,

- $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2, X_3, X_4 \sim \chi_1^2$  for the regular observations.
- $X_1 \sim \mathcal{N}(0, 3) + C_D$  and  $X_2, X_3, X_4 \sim \chi_4^2 + C_D$  for the contaminated observations.

In CASE 2 dependence between the variables  $X_i$  is introduced as follows.

- $X_1 \sim \mathcal{N}(0, 1)$  and  $X_2, X_3, X_4 \sim 1/(X_1^2 + 1)^2 + \sqrt{\chi_1^2}$  for the non-contaminated subsample (with  $\chi_1^2$  independent of  $X_1$ ),
- $X_1 \sim \mathcal{N}(0, 3) + C_D$  and  $X_2, X_3, X_4 \sim 1/(X_1^2 + 1)^2 + \sqrt{\chi_1^2} + C_D$  for the contaminated subsample (with  $\chi_1^2$  independent of  $X_1$ ).

Both the fuzzy M-estimate of location and the fuzzy MDD-based M-estimate of location using the Hampel loss function are computed for each contaminated sample using an algorithm as in [10]. Their population values are approximated by Monte Carlo with 10000 iterations and the performance of both proposals is compared in terms of the corresponding mean square error also approximated by Monte Carlo with 1000 iterations.

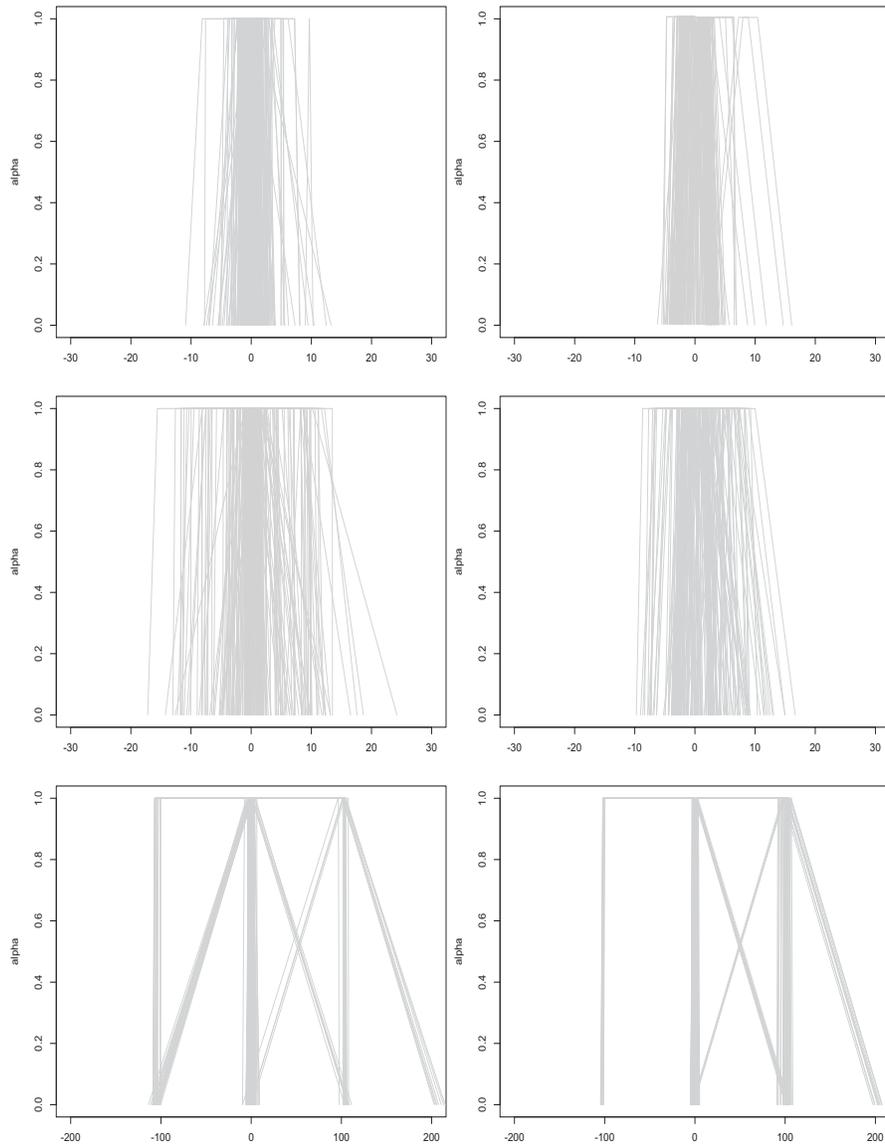
## 4.1 Results

Table 4.1 contains the results of the comparative analysis. In order to understand these results better, Figure 1 shows some samples generated using the procedure explained in this section with some illustrative choices of the contamination parameters  $c_p$  and  $C_D$ . It can be concluded that

**Table 1** Empirical comparison of the fuzzy M-estimate of location (Hampel) and the fuzzy MDD-based M-estimate of location (MDD-H) using the Hampel loss function

$c_p$	$C_D$	CASE 1		CASE 2	
		Hampel	MDD-H	Hampel	MDD-H
0	0	0.03157	<b>0.02077</b>	0.39544	<b>0.11508</b>
0.1	0	0.03242	<b>0.02790</b>	0.38452	<b>0.11492</b>
0.1	1	<b>0.03272</b>	0.03460	0.40867	<b>0.13536</b>
0.1	5	<b>0.03079</b>	0.04271	0.36526	<b>0.18030</b>
0.1	10	<b>0.02903</b>	0.02995	0.35188	<b>0.14413</b>
0.1	100	0.02977	<b>0.02258</b>	0.35691	<b>0.11714</b>
0.2	0	<b>0.04257</b>	0.05940	0.37045	<b>0.11383</b>
0.2	1	<b>0.04595</b>	0.08613	0.43351	<b>0.15930</b>
0.2	5	<b>0.04685</b>	0.16308	0.36094	<b>0.33573</b>
0.2	10	<b>0.04444</b>	0.08832	0.32239	<b>0.23062</b>
0.2	100	0.04649	<b>0.02491</b>	0.31370	<b>0.11416</b>
0.4	0	<b>0.10742</b>	0.20288	0.34975	<b>0.11612</b>
0.4	1	<b>0.15262</b>	0.37964	0.50878	<b>0.23767</b>
0.4	5	<b>0.35651</b>	1.82037	<b>0.69800</b>	1.55067
0.4	10	<b>0.54289</b>	3.21184	<b>0.87081</b>	2.37917
0.4	100	0.70982	<b>0.03470</b>	0.97474	<b>0.11623</b>

- there is no uniformly best estimator, since both the fuzzy M-estimator of location and the fuzzy MDD-based M-estimator of location can provide us with the most convenient estimate depending on the analyzed situation.
- the distribution seems to have more influence than the amount of contamination on the choice of the best estimator.
- the larger the difference between the outlier and the non-contaminated data, the more the fuzzy MDD-based M-estimate of location seems to improve the results of the fuzzy M-estimate of location, since even after scaling the observations, the difference between the outlier and the rest of data is still clear and the Hampel loss function can conveniently cope with it. This can be clearly noticed when  $c_p = 0.4$  and  $C_D = 100$  (graphics at the bottom in Figure 1). Due to the chosen distributions, the opposite is shown when  $c_p = 0.4$  and  $C_D = 5$  (graphics at the middle in Figure 1), since outliers do not lie so far away from the non-contaminated part of the sample in both CASE 1 and CASE 2 and adjusting the tuning parameters in the Hampel loss function seems to be a better option than scaling all the observations (and therefore shortening the distances between outliers and the rest of data even more). However, when the contamination proportion  $c_p$  decreases to 0.1 and  $C_D$  remains equal to 5 (graphics at the top in Figure 1), it can be seen that CASES 1 and 2 are not so similar as in the previous situations, and the larger distance between outliers and non-contaminated data in CASE 2 than in CASE 1 makes the fuzzy MDD-based M-estimator of location improve the behaviour of the fuzzy M-estimator of location only in CASE 2.



**Fig. 1** A sample of generated fuzzy numbers from CASE 1 (left column) and CASE 2 (right column) is compared when  $(c_p, C_D)$  is chosen to be  $(0.1, 5)$  –top–,  $(0.4, 5)$  –middle– and  $(0.4, 100)$  –bottom–

## 5 Concluding remarks

In this paper, M-estimators of location with unknown dispersion have been extended to the fuzzy number-valued settings in order to provide a scale

equivariant alternative to fuzzy M-estimators of location. They have been defined in terms of the median ( $D_{\theta}^{\varphi}$ -)distance deviation about the 1-norm median. The two alternatives have been empirically compared and, even when there is no uniformly best estimator, it seems that the fuzzy MDD-based M-estimator of location may provide us with the best results when the distance between the outliers and the ‘standard’ data is large enough. Therefore, it could better identify which observations should be considered outliers indeed since it is using the information of the global dispersion. However, it would be advisable to complete this simulation study in the future to deal with other kinds of distributions and check whether these preliminary conclusions would remain.

## Acknowledgments

This research has been partially supported by the Spanish Ministry of Economy and Competitiveness through the Grant MTM2013-44212-P and the Principality of Asturias/FEDER Grant GRUPIN14-101. Their support is gratefully acknowledged.

## References

1. Bertoluzza C, Corral N, Salas A (1995) On a new class of distances between fuzzy numbers. *Math & Soft Comput* 2:71–84
2. Blanco-Fernández A, Casals RM, Colubi A, Corral N, García-Bárzana M, Gil MA, González-Rodríguez G, López MT, Lubiano MA, Montenegro M, Ramos-Guajardo AB, de la Rosa de Sáa S, Sinova B (2014) A distance-based statistical analysis of fuzzy number-valued data. *Int J Approx Reason* 55:1487–1501
3. De la Rosa de Sáa S, Gil MA, González-Rodríguez G, López MT, Lubiano MA (2015) Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Trans Fuzzy Syst* 23(1):111–126
4. De la Rosa de Sáa S, Lubiano MA, Sinova B, Filzmoser (2015) Robust scale estimators for fuzzy data. *Adv Data Anal Class*. Accepted, doi:10.1007/s11634-015-0210-1
5. Hampel FR (1974) The influence curve and its role in robust estimation. *J Amer Stat Assoc* 69:383–393
6. Kim JS, Scott CD (2012) Robust kernel density estimation. *J Mach Learn Res* 13:2529–2565
7. Montenegro M, Casals MR, Lubiano MA, Gil MA (2001) Two-sample hypothesis tests of means of a fuzzy random variable. *Inf Sci* 133:89–100
8. Puri ML, Ralescu DA (1986) Fuzzy random variables. *J Math Anal Appl* 114:409–422
9. Sinova B, Van Aelst S (2017) Tukey’s Biweight Loss Function for Fuzzy Set-Valued M-estimators of Location. In: Ferraro MB, Giordani P, Vantaggi B, Gagolewski M, Gil MA, Grzegorzewski P, Hryniewicz O (eds) *Soft Methods for Data Science. Advances in Intelligent Systems and Computing*. Springer, Switzerland
10. Sinova B, Gil MA, Van Aelst S (2016) M-estimates of location for the robust central tendency of fuzzy data. *IEEE Trans Fuzzy Syst* 24(4):945–956

11. Sinova B, Pérez-Fernández S, Montenegro M (2015) The Wabl/Ldev/Rdev Median of a Random Fuzzy Number and Statistical Properties. In: Grzegorzewski P, Gagolewski M, Hryniewicz O, Gil MA (eds) Strengthening Links Between Data Analysis and Soft Computing. Advances in Intelligent Systems and Computing. Springer, Heidelberg
12. Sinova B, Gil MA, Colubi A, Van Aelst S (2012) The median of a random fuzzy number. The 1-norm distance approach. *Fuzzy Sets Syst* 200:99–115