www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

ADVANCES IN GEODESY AND GEOINFORMATION
(formerly GEODESY AND CARTOGRAPHY)
Vol. 71, no. 1, article no. e17, 2022

https://doi.org/10.24425/gac.2022.141170

Original paper

Part of Topical collection:
"Advancements in Applied Geoinformatics"

Absolute $\mathrm{M_{split}}$ estimation as an alternative for robust M-estimation

Robert Duchnowski*, **Patrycja Wyszkowska**

University of Warmia and Mazury, Olsztyn, Poland,
e-mail: robert.duchnowski@uwm.edu.pl; ORCID: http://orcid.org/0000-0002-6331-8345
e-mail: patrycja.wyszkowska@uwm.edu.pl; ORCID: http://orcid.org/0000-0002-0080-7897

*Corresponding author: Robert Duchnowski, e-mail: robert.duchnowski@uwm.edu.pl

Received: 2021-01-24 / Accepted: 2022-03-21

Abstract: The problem of outlying observations is very well-known in the surveying data processing. Outliers might have several sources, different magnitudes, and shares within the whole observation set. It means that it is not possible to propose one universal method to deal with such observations. There are two general approaches in such a context: data cleaning or robust estimation. For example, the robust M-estimation has found many practical applications. However, there are other options, such as R-estimation or the absolute $\mathrm{M_{split}}$ estimation. The latter method was created to be less sensitive to outliers than the squared $\mathrm{M_{split}}$ estimation (the basic variant of $\mathrm{M_{split}}$ estimation). From the theoretical point of view, the absolute $\mathrm{M_{split}}$ estimation cannot be classified as a robust method; however, it was proved that it could be used in such a context under certain conditions. The paper presents the primary comparison between that method and a conventional robust M-estimation. The results show that the absolute $\mathrm{M_{split}}$ estimation predominates over the classical methods, especially when the percentage of outliers is high. Thus, that method might be used to process LiDAR data, including mismeasured points. Processing synthetic data from terrestrial laser scanning or airborne laser scanning confirms that the absolute $\mathrm{M_{split}}$ estimation can deal with outliers sufficiently.

Keywords: laser scanning, robust estimation, M-estimation, absolute $\mathrm{M_{split}}$ estimation

1. Introduction

The problem of outlying observations is well-known in the geodetic data processing. It is no doubt that outliers might disturb the estimation of the parameters of the observation models. It is essential to realize that outlying observations might occur, and their sources can be different. Usually, we simply define the concept of an outlier as an unknowingly occurring gross error in an observation. However, in some modern measurement techniques such as GNSS positioning and LiDAR, outliers might result from their own specific mostly known characteristics (e.g., Baselga and García-Asenjo, 2008; Carrilho et al., 2018). In the latter groups of techniques, including airborne laser scanning (ALS) or terrestrial laser scanning (TLS), some points might be "mismeasured." Namely, such points might not concern the surface we are interested in but other structures or obstacles (e.g. Carrilho and Galo, 2019; Wyszkowska et al., 2021). Several approaches or methods have been elaborated to cope with outlying observations, considering their different sources, magnitudes, or percentages of outliers in the observation set. The first approach is based on the application of several statistical tests to flag outliers and finally, remove them from the observation set (e.g., Baarda, 1968; Pope, 1976; Lehmann, 2013a, 2013b; Hekimoglu et al., 2015; Duchnowski and Wyszkowska, 2022). In such a case, we just ignore the observations suspected to be affected by gross errors. The second approach to outliers is the application of the robust estimation methods. One should list several methods that belong to the class of robust M-estimations concerning geodetic applications. The most popular methods are the Huber method, the Tukey method, the least absolute deviation method (LAD), the Danish method, IGG scheme (Xu, 1989; Yang, 1994; Hekimoglu and Berber, 2003; Chang and Guo, 2005; Wyszkowska and Duchnowski, 2020b). In some applications, one can also consider R-estimation usage (Kargoll, 2005; Duchnowski, 2013). There are also other possible methods, which are not listed above. Here one could mention a modern estimation method called M_{split} estimation introduced in Wisniewski (2009). Generally, such a method was developed to process heterogeneous observation sets, not to deal with outliers. However, it could also be applied in selected problems as a robust estimation method thanks to its general features (Li et al., 2013; Duchnowski and Wisniewski, 2020). In such a case, the method is often modified by applying the additional (virtual) functional model that is meant to "collect" outliers (Zienkiewicz, 2015). Such an approach can provide correct results, but under some specific assumptions concerning location of outliers, hence, it is not universal (Wyszkowska, 2021). The problem is that the basic M_{split} estimation variant, namely the squared M_{split} estimation, cannot be classified as a method robust against outliers (Duchnowski and Wisniewski, 2019, 2020). Thus, this paper focuses on another M_{split} estimation variant, namely the absolute M_{split} estimation, in which the influence functions are derived from the influence function of LAD method. The empirical tests showed that it made the estimates robust to gross errors of a moderate magnitude (Wyszkowska and Duchnowski, 2019). Thus, the main aim of the paper is to compare results of the absolute M_{split} estimation with basic M-estimations, and to show its applicability in the processing observation sets including outliers, especially TLS measurements.

The paper is organized in the following way: Section 2 introduces the methods applied,

Section 3 presents basic empirical tests of robustness in the case of the univariate model, Section 4 shows processing the simulated TLS data in determining one or two surfaces. Section 5 sums up the tests and conclusions.

2. Methods and general assumptions

Let us consider the conventional functional model of geodetic observations in the following linear form:

$$\mathbf{y} = \mathbf{A}\mathbf{X} + \mathbf{v}, \tag{1}$$

where: $\mathbf{y} = [y_1, \ldots, y_n]^T$ is an observation vector, $\mathbf{v} = [v_1, \ldots, v_n]^T$ is a vector of random errors, $\mathbf{X} = [X_1, \ldots, X_m]^T$ is a parameter vector, $\mathbf{A}$ is a known rectangular matrix of size $n \times m$; here, it is assumed to be full column rank. We assume that the expected value $E(\mathbf{y}) = \mathbf{A}\,\mathbf{X}$ and all observations are of the same accuracy. Hence, the weight matrix of observations is equal to the identity matrix ($\mathbf{P} = \mathbf{I}$). Then, the least squares (LS) estimate of the parameter vector is given as:

$$\hat{\mathbf{X}}_{LS} = \left(\mathbf{A}^T\,\mathbf{A}\right)^{-1}\mathbf{A}^T\,\mathbf{y}. \tag{2}$$

A similar formula could apply to M-estimation leading to the following form of M-estimator of the parameter vector:

$$\hat{\mathbf{X}}_{M} = \left(\mathbf{A}^T\,\mathbf{W}\,\mathbf{A}\right)^{-1}\mathbf{A}^T\,\mathbf{W}\,\mathbf{y}, \tag{3}$$

where: $\mathbf{W}$ is a diagonal matrix of the weights for which $[\mathbf{W}]_{ii} = w\,(\hat{v}_i)$, $w\,(\hat{v}_i)$ is a respective weight function related to the particular M-estimation variant, $\hat{v}_i$ is the standardized error of the ith observation, $[\circ]_{ii}$ is the ith diagonal element of the matrix. The respective M-estimates are computed in the iterative process, which ends when the parameter vector is no longer changing between the iteration steps. Such a solution refers to the iteratively reweighted least squares (IRLS); however, other solutions of the optimization problem, such as the Newton-Raphson method, are also possible (e.g. Amiri-Simkooei, 2003; Wisniewski, 2009; Wyszkowska and Duchnowski, 2019).

Several different weight functions can be applied in Eq. (3). Here, we consider two weight functions that are popular in geodetic applications and differ in general features. The first weight function refers to the Huber method and can be written as follows (e.g. Gui and Zhang, 1998; Baselga, 2007; Ge et al., 2013):

$$w_H\,(\hat{v}_i) = \begin{cases} 1 & \text{for} \quad |\hat{v}_i| \le a, \\ \dfrac{a}{|\hat{v}_i|} & \text{for} \quad |\hat{v}_i| > a, \end{cases} \tag{4}$$

where: a is a positive constant (usually assumed between 1.5 and 3.5). We also consider the Tukey weight function, which has the rejection points in contrast to the Huber weight function. It can be given as (Gui and Zhang, 1998; Ge et al., 2013):

$$w_T\,(\hat{v}_i) = \begin{cases} \left(1 - \dfrac{|\hat{v}_i|^2}{a^2}\right)^2 & \text{for} \quad |\hat{v}_i| \le a, \\ 0 & \text{for} \quad |\hat{v}_i| > a, \end{cases} \tag{5}$$

where this time, the constant a is generally lower than or equal to 6.

In the case of $\mathrm{M_{split}}$ estimation, the traditional functional model of Eq. (1) should be split into two competitive models (Wisniewski, 2009):

$$\mathbf{y} = \mathbf{A}\,\mathbf{X} + \mathbf{v} \quad \Rightarrow \quad \begin{cases} \mathbf{y} = \mathbf{A}\,\mathbf{X}_{(1)} + \mathbf{v}_{(1)}\,, \\ \mathbf{y} = \mathbf{A}\,\mathbf{X}_{(2)} + \mathbf{v}_{(2)}\,, \end{cases} \tag{6}$$

where: $\mathbf{X}_{(1)}$ and $\mathbf{X}_{(2)}$ are the competitive versions of the parameter vector $\mathbf{X}$, $\mathbf{v}_{(1)}$ and $\mathbf{v}_{(2)}$ are the competitive versions of the vector of random errors $\mathbf{v}$. Note that it is also possible to split the conventional model into q competitive ones, which lead to $\mathrm{M_{split}}_{(q)}$ estimation (Wisniewski, 2010). Considering the absolute $\mathrm{M_{split}}$ estimation (here denoted as AMS estimation), the competitive versions of parameters are determined by the Newton-Raphson method in the following iterative process (Wyszkowska and Duchnowski, 2019):

$$\begin{aligned} \mathbf{X}_{(1)}^{j} &= \mathbf{X}_{(1)}^{j-1} + d\mathbf{X}_{(1)}^{j} = \mathbf{X}_{(1)}^{j-1} - \left[\mathbf{H}_{(1)}\left(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}\right)\right]^{-1} \mathbf{g}_{(1)}\left(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}\right), \\ \mathbf{X}_{(2)}^{j} &= \mathbf{X}_{(2)}^{j-1} + d\mathbf{X}_{(2)}^{j} = \mathbf{X}_{(2)}^{j-1} - \left[\mathbf{H}_{(2)}\left(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}\right)\right]^{-1} \mathbf{g}_{(2)}\left(\mathbf{X}_{(1)}^{j-1}, \mathbf{X}_{(2)}^{j-1}\right), \end{aligned} \tag{7}$$

where: $d\mathbf{X}_{(l)}$ is an increment to parameter vector, $\mathbf{H}_{(l)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ are the Hessians, $\mathbf{g}_{(l)}(\mathbf{X}_{(1)}, \mathbf{X}_{(2)})$ are the gradients, l is equal 1 or 2. The Hessians and gradients are defined in the following way:

$$\begin{aligned} \mathbf{H}_{(1)}\left(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\right) &= 2\mathbf{A}^{T}\mathbf{w}_{(1)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right)\mathbf{A}, \\ \mathbf{H}_{(2)}\left(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\right) &= 2\mathbf{A}^{T}\mathbf{w}_{(2)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right)\mathbf{A}, \end{aligned} \tag{8}$$

$$\begin{aligned} \mathbf{g}_{(1)}\left(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\right) &= -2\mathbf{A}^{T}\mathbf{w}_{(1)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right)\mathbf{v}_{(1)}\,, \\ \mathbf{g}_{(2)}\left(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\right) &= -2\mathbf{A}^{T}\mathbf{w}_{(2)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right)\mathbf{v}_{(2)}\,. \end{aligned} \tag{9}$$

The matrices $\mathbf{w}_{(l)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right)$ are computed by applying the respective weight functions:

$$\begin{aligned} \mathbf{w}_{(1)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right) &= \mathrm{diag}\left[w_{(1)}\left(v_{1(1)}, v_{1(2)}\right), \ldots, w_{(1)}\left(v_{n(1)}, v_{n(2)}\right)\right], \\ \mathbf{w}_{(2)}\left(\mathbf{v}_{(1)}, \mathbf{v}_{(2)}\right) &= \mathrm{diag}\left[w_{(2)}\left(v_{1(1)}, v_{1(2)}\right), \ldots, w_{(2)}\left(v_{n(1)}, v_{n(2)}\right)\right], \end{aligned} \tag{10}$$

where: $\mathrm{diag}(\circ)$ is a diagonal matrix. The weight functions of AMS estimation are derived in the following forms (Wyszkowska and Duchnowski, 2019):

$$\begin{aligned} w_{(1)}\left(v_{i(1)}, v_{i(2)}\right) &= \begin{cases} \dfrac{\left|v_{i(2)}\right|}{2c} & \text{for} \quad \left|v_{i(1)}\right| < c, \\ \dfrac{\left|v_{i(2)}\right|}{2\left|v_{i(1)}\right|} & \text{for} \quad \left|v_{i(1)}\right| \geq c, \end{cases} \\ w_{(2)}\left(v_{i(1)}, v_{i(2)}\right) &= \begin{cases} \dfrac{\left|v_{i(1)}\right|}{2c} & \text{for} \quad \left|v_{i(2)}\right| < c, \\ \dfrac{\left|v_{i(1)}\right|}{2\left|v_{i(2)}\right|} & \text{for} \quad \left|v_{i(2)}\right| \geq c, \end{cases} \end{aligned} \tag{11}$$

where: c is an assumed small positive constant. The presented iterative process called the parallel process and the conventional iterative process, convenient for the squared M_{split} estimation, were addressed in detail in the paper (Wyszkowska and Duchnowski, 2020a).

3. Basic empirical tests

First, let us concern the case of the univariate model; hence, let $\mathbf{X} = X$, $\mathbf{X}_{(1)} = X_{(1)}$, $\mathbf{X}_{(2)} = X_{(2)}$. Here, we assume that $X = X_{(1)} = 0$ and $X_{(2)} = 5$. Let the observation vector be simulated assuming that observation errors are normally distributed for $N(0, 1)$. There are five observations concerning the location parameter $X_{(1)}$ and a different number of observations concerning the location parameter $X_{(2)}$: one such an observation in Variant I, two in Variant II, three in Variant III, four in Variant IV, and five observations in Variant V. Note that in the classical M-estimation, the latter observation group should be considered as outliers. In the case of AMS estimation, one has a different attitude. We assume *a priori* that there are two observation groups whose location parameters are estimated. In the paper context, from the two solutions, $\hat{X}_{(1)}$ and $\hat{X}_{(2)}$, we are interested only in the first one. In this section, we neglect the estimate of $X_{(2)}$ as describing the location of outliers, which is out of our interest.

Now, let the location parameter $X_{(1)}$ be estimated by applying LS, M-estimation methods, or AMS estimation. We use the following steering parameters: $a = 2$ for the Huber method, $a = 6$ for the Tukey method, and $c = 0.001$ for AMS estimation. To describe how the outlying observations influence the estimation results, we simulate 1000 observation vectors. It allows us to compute the Monte Carlo estimate of the location parameters (as the mean value) and assess the estimation accuracy by calculating the root-mean-square deviation (*RMSD*):

$$\text{Mean}\left(\hat{X}\right) = \frac{\sum_{i=1}^{1000} \hat{X}_i}{1000}, \qquad RMSD\left(\hat{X}\right) = \sqrt{\frac{\sum_{i=1}^{1000} \left(\hat{X}_i - X\right)^2}{1000}}, \tag{12}$$

where: $\hat{X}_i$ is the parameter estimate obtained in ith simulation.

The results are presented in Figure 1. There is no doubt that the outlying observations disturb estimates of LS and M-estimations. Only in Variant I, the results of the robust methods are significantly better than the results of LS estimation. The differences between estimates in question are very small in the other variants, and the effects are unacceptable. It stems from the fact that the magnitude of the gross errors is not big enough to make the outliers be detected "easily," especially when the number of outliers is bigger (see Variants III–V). In contrast, the results of AMS estimation are similar to each other for all the variants. They are always better than the results of the other methods, and AMS estimation predominates over other methods for a growing number of outliers.

The observation sets simulated in the variants at hand seem natural to be processed by applying M_{split} estimation with two competitive split functional models as in Eq. (6).

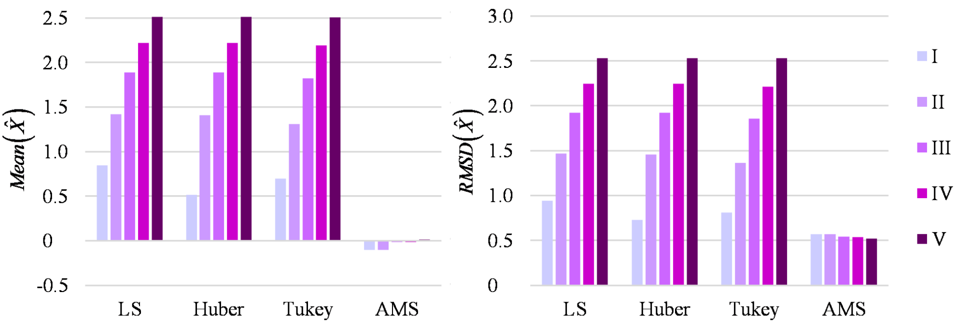

Fig. 1. Means and *RMSD*s for Variants I–V

This is why AMS estimation provides better results than the conventional methods. Let us now supplement the observation set with an outlier that does not coincide with two groups of the observations. In each variant, we add one outlying observation simulated from the uniform distribution in the interval $\langle -5, -1 \rangle$. Hence, this outlier lies on the "opposite" side to the second group of observations. Figure 2 presents the results for all new variants denoted as Variants VI–X.

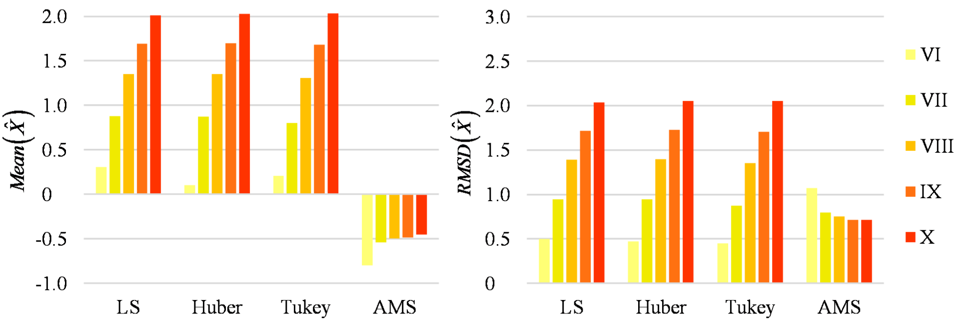

Fig. 2. Means and *RMSD*s for Variants VI–X

The additional outlier deteriorates the success of AMS estimation, especially when we consider the mean value. The results in Variant VI seem the most affected by such an outlier. Note that we have one group of five observations in this variant and two outliers lying on the opposite sides of that group. Thus, the AMS estimation algorithm has a problem with dividing observations into two groups. In Variants VII-X, we have more observations close to 5; hence, the method has less of a problem with the correct placement of the location parameter estimates. In those variants, AMS estimation provides better results than the classical methods. Note that the better results of the conventional methods in Variants VI-X than in respective Variants I-V result from the placement of the additional outlier, not from the feature of the methods in question.

The simple tests presented here show that AMS estimation might predominate over the classical robust M-estimation in the case of a single outlier or when the outliers lie on the same side of the main group of observations. This conclusion is especially important for processing TLS data when usually most of the outlying observations are so-called positive, lying above the surface we are interested in (Carrilho et al., 2018; Wyszkowska et al., 2021).

4. Processing simulated TLS data

This section focuses on processing LiDAR data. We consider two cases; in the first one, we process simulated results of scanning a single surface, in the second case – two surfaces.

Let us consider a terrain profile determined from TLS data by approximation of polynomials. Of course, it seems that such an assumption is very basic; however, it has found a practical application. The polynomial approximation of the surface measured using the TLS technique was used to determine the beam deformation (Cabaleiro et al., 2015). A similar approach was applied to determine terrain profiles from the real TLS data (Wyszkowska et al., 2021). Thus, examining such an approach seems to be advisable from the practical point of view. For the purposes of the paper, we can assume that the terrain profile of 20 m-long can be approximated well by the polynomial of the second degree (the theoretical coefficients of the polynomial are accepted as: $c_2 = 0.003$, $c_1 = -0.04$, $c_0 = 1$). It is the base for simulating 100 observations, heights H_i at the randomly chosen distances $D_i \in \langle 0 \text{ m}, 20 \text{ m}\rangle$. The observation errors are assumed to be normally distributed with the expected value of 0 mm and the standard deviation of 2 mm. We can also imagine that some observations might be affected by gross errors ranging from +10 to +100 mm. Such outliers simulate mismeasured points, namely measurements of vegetation cover (shrubs, grass, fallen leaves) or other obstacles in the TLS data. A detailed description of similar simulations can be found in (Wyszkowska et al., 2021). The simulated observation sets are presented in Figure 3.

The terrain profile P is approximated by applying the estimated polynomial coefficients. Let them be estimated by the chosen in the previous section M-estimation variants, the Huber and the Tukey methods, and AMS estimation (the values of the steering parameters remain the same). Note that we consider the split functional model in the case of AMS estimation; thus, we obtain two competitive solutions. Here, we choose only one solution, the polynomial whose graph is located lower (Wisniewski, 2009; Wyszkowska et al., 2021). The general results are presented in panel a) of Figure 4; panel b) presents the estimated profiles for the chosen interval from 5 to 10 m to underline the difference between estimation methods. The best results are obtained for AMS estimation, especially for the growing percentage of outliers.

To assess the accuracy of the estimated profiles, we use $RMSD\left(\hat{H}\right)$ in the following form (Wyszkowska et al., 2021)

$$RMSD\left(\hat{H}\right) = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(\hat{H}_i - H_i\right)^2}{n}}, \tag{13}$$

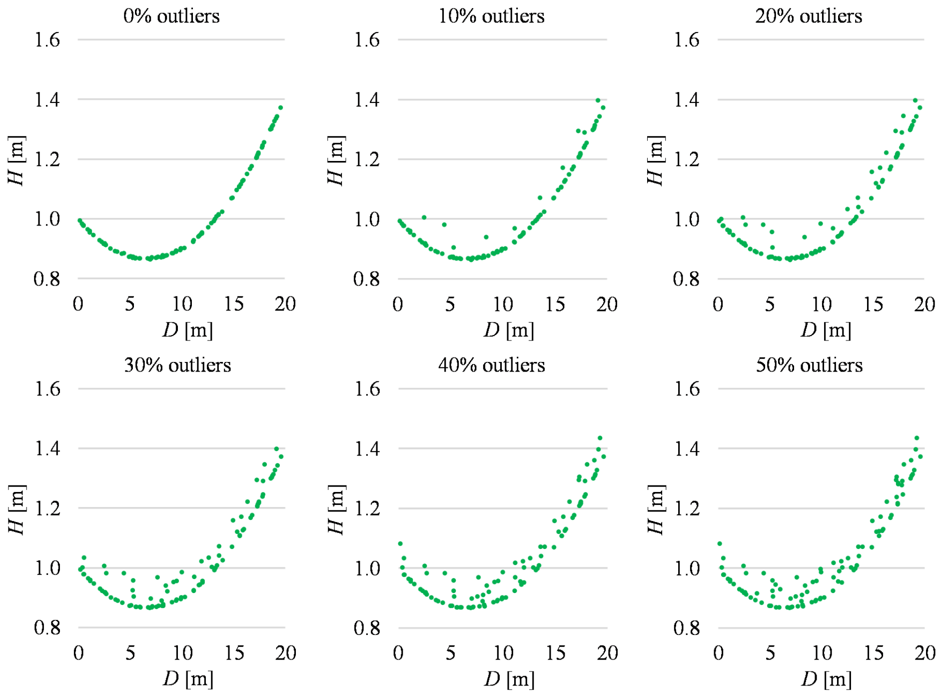

Fig. 3. The simulated observation sets

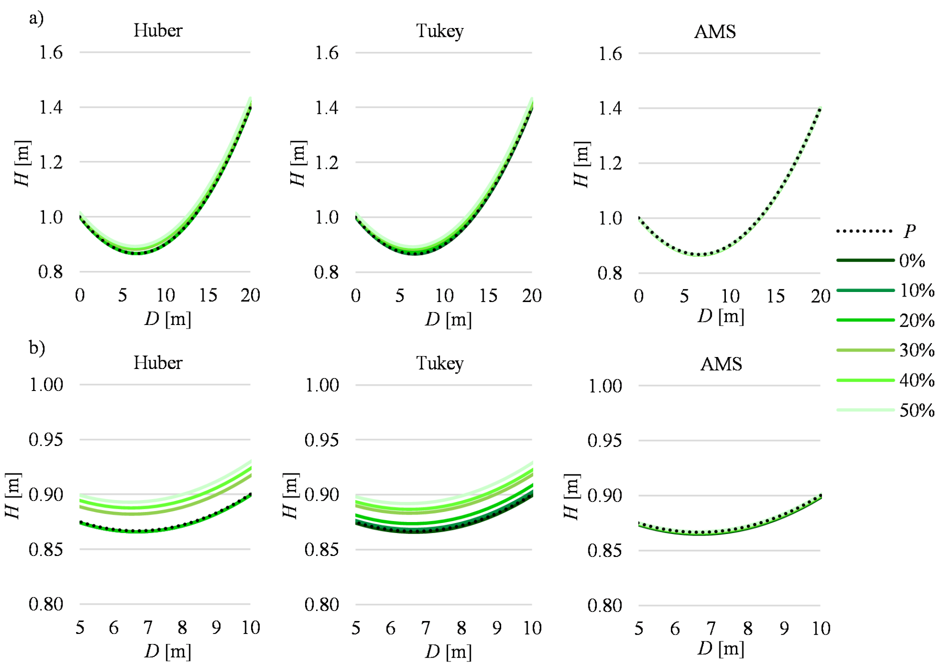

Fig. 4. Estimated terrain profiles for the Huber, Tukey, and AMS estimates for different percentages of outliers: for the interval from 0 to 20 m, for the interval from 5 to 10 m (P is the simulated terrain profile)

where: $\hat{H}_i$ are the estimated heights, H_i are the simulated heights. Here $n = 41$, which is the number of points for which heights are calculated for distances $D_j = j \cdot 0.5$ m ($j = 0, \ldots, 40$).

Figure 5 shows $RMSD\left(\hat{H}\right)$ determined for all simulated observation sets and three methods mentioned. M-estimates predominate over AMS method if the observation set is free from outliers. The accuracy of the Tukey method decreases rapidly with the growing percentage of outliers. The Huber method and AMS estimation have similar accuracy up to 30% of outliers; for more outliers, AMS estimation predominates. It is also worth noting that the accuracy of AMS estimates increases slightly with the growing percentage of outliers, which is not such an apparent feature of the method. However, the explanation is natural. If there is a higher share of outliers, then the method can distinguish that observation group from the group of regular observations easier.

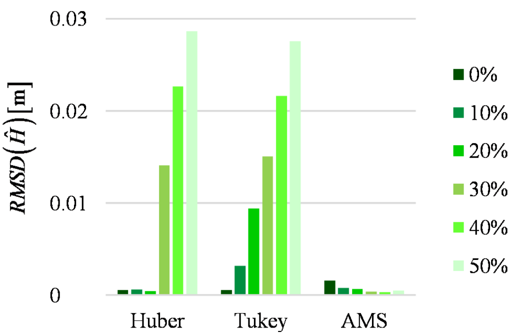

Fig. 5. $RMSD\left(\hat{H}\right)$ for the Huber, Tukey, and AMS estimates for different percentages of outliers

In the following numerical example, we consider LiDAR data that refer to two surfaces. Such an observation set might result from, for example, measuring the terrain (denoted P) and the vegetation cover (denoted P'). Let us assume that measurements concern the profile of 50 m long. The terrain profile is described by the polynomial of the first degree ($c_1 = 0.035$, $c_0 = 1$), and the vegetation cover by the polynomial of the second degree ($c_2 = -0.001$, $c_1 = 0.08$, $c_0 = 1.4$). We simulate randomly 100 observations, heights H_i at the randomly chosen distance $D_i \in \langle 0 \text{ m}, 50 \text{ m}\rangle$. The observation errors are assumed to be normally distributed with the expected value of 0 mm and the standard deviation of 50 mm, which are acceptable in the application in question (e.g. Crespo-Peremarch et al., 2018). We also assume that data might randomly be disturbed by gross errors from 0.20 m to 0.80 m (if such errors concern the terrain profile – outliers are considered positive, but if gross errors involve the vegetation cover – outliers are negative). In the case of AMS estimation, one uses the whole observation set; however, when applying the classical methods, one should consider two sets separately – observations of the terrain profile and observations of the vegetation cover, respectively. The problem is that we do not know the actual data division (the real assignment of each observation to either of the subsets). Thus, we decided to create the subsets in the following way: we consider ten intervals of 5 m long; in each interval, the largestvalued observations, accounting the half of all, are assigned as measurements of the vegetation cover, whereas the other observations as measurements of the terrain profile. The simulated observation sets are presented in Figure 6. Like in the previous example, the polynomial coefficients are now estimated by applying the Huber and the Tukey methods and AMS estimation (the values of the steering parameters remain the same). The estimated profiles are presented in Figure 7. Generally, all methods estimate the vegetation cover correctly; however, the advantage of AMS estimation application arises while increasing the number of outlying observations. Almost the same conclusion concerns the estimated terrain profiles.

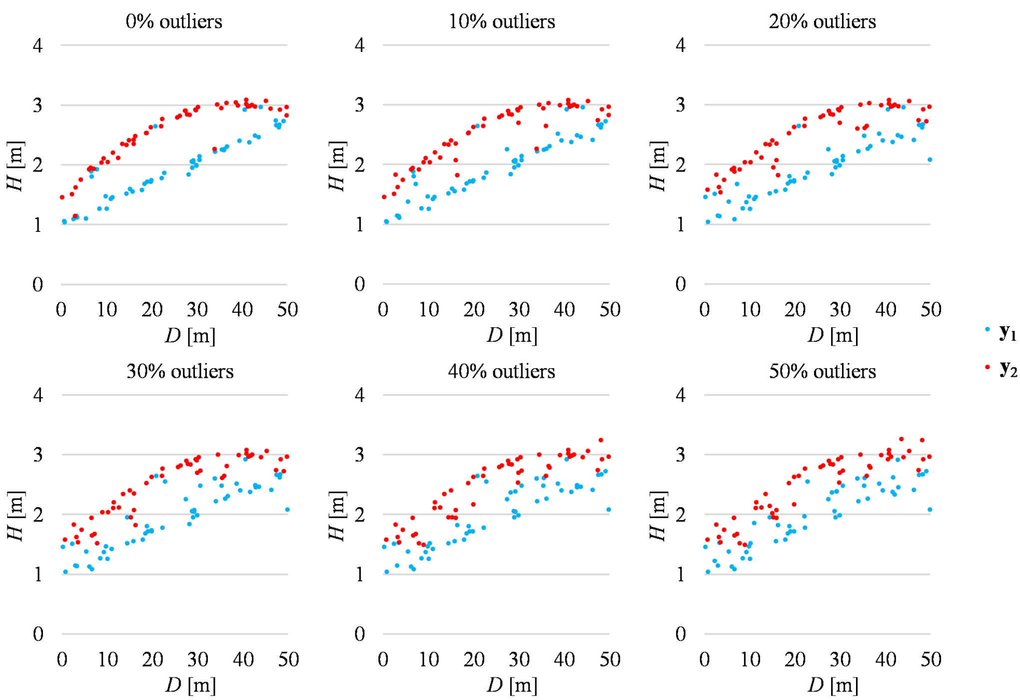

Fig. 6. The simulated observation sets (red – observations assigned to the vegetation cover; blue – to the terrain profile)

One can say that the biggest problem concerns estimation of the vegetation cover and the terrain profile at the end of the interval under consideration (where D is close to 50 m). It stems from the fact that both estimated lines are tight together, and the occurrence of outliers makes the subset separation even harder. To compare the results of different estimation methods, we also compute the accuracy of the fit of all estimated profiles by applying Eq. (13). The respective *RMSD*s are calculated like in the previous example, but by taking $n = 101$; they are presented in Figure 8. The Huber method provides the most accurate results when outliers account for no more than 10%; the effects of AMS estimation are only a little bit worse in such a case. For the growing percentage of outliers, AMS estimation predominates. It shows that conventional methods cannot cope

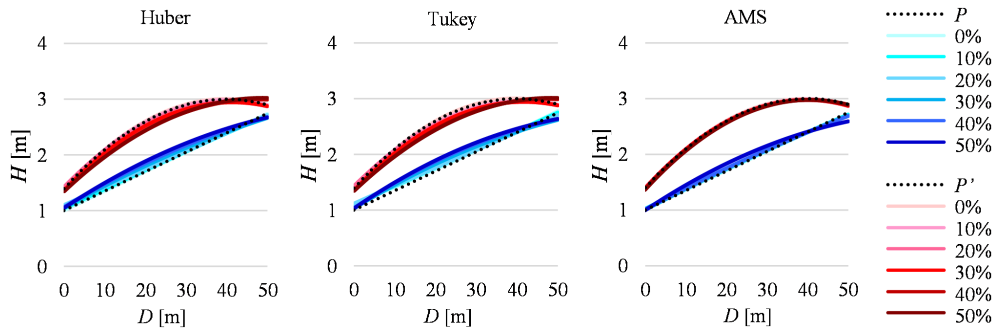

Fig. 7. Estimated terrain profiles for Huber, Tukey, and AMS estimates for different percentages of outliers (P is the simulated terrain profile, P' is the simulated vegetation cover profile)

with a higher share of outliers. One should also realize that the conventional methods process the separated subsets, whereas AMS estimation uses no *a priori* information about observation assignment. Such an assignment is done during the estimation process. That numeric example proved that approach as better and more efficient than the classical one.

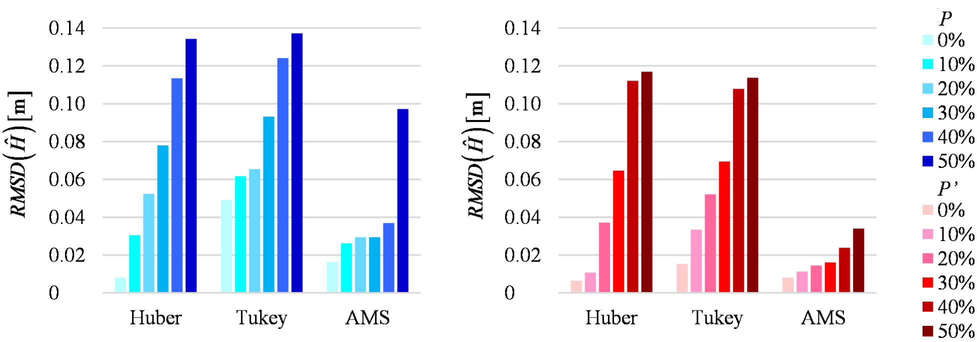

Fig. 8. $RMSD\left(\hat{H}\right)$ for Huber, Tukey, and AMS estimates for different percentages of outliers (P is the simulated terrain profile, P' is the simulated vegetation cover profile)

5. Conclusions

The absolute $\mathrm{M}_{\text{split}}$ estimation is a novel method of processing geodetic observations. It was created to process heterogeneous observation sets (usually, we assume that we deal with two subsets; however, it is possible to take a higher number of observation groups). The general features of the method allow us to use it for other purposes. That paper focuses on applying AMS estimation as a robust estimation method. The previous research showed that AMS estimation could deal with gross errors of a moderate magnitude; however, the method cannot be classified as robust against outliers in general.

The numerical tests performed for the univariate model show that AMS estimation provides better results than the classical robust method belonging to the class of M-estimation if only outliers lie close to one another. In such a case, the magnitude of gross errors is irrelevant. If outliers are not grouped, the results of AMS estimation worsen; however, the method itself is still a good alternative for the classical robust methods. It is especially evident for the growing number of outliers, where the accuracy of AMS estimation is much higher than the accuracy of M-estimation. The iterative process of AMS estimation is more complicated than the classical process of M-estimation, and it requires more time. However, the difference in processing time has no practical significance.

The paper shows that AMS estimation can be used for processing LiDAR data. The numerical examples concern approximation of one surface or two surfaces from one TLS or ALS data. The difference between those cases is significant. In the first example, we have an observation set that is "natural" for AMS estimation – there are two observation groups: regular observations and outliers. In the second case, we have two groups of regular observations and additional outliers, making the estimation more difficult. Both examples confirmed that AMS estimation is a good alternative for other robust methods. The case of one surface estimation illuminates not an obvious feature of M_{split} estimation. For the growing percentage of outliers, the accuracy of AMS estimation increases. It stems from the fact that it is easier to distinguish two groups of observations when the number of outliers is high. Hence, it is also easier to estimate parameters within the split functional model. That example also shows that AMS estimation predominates over M-estimation when the percentage of outliers is higher than 20%. In the second example (ALS data processing), AMS estimation provides better results when the share of outliers is equal to or higher than 10%. However, we should realize that we simulate outlying observations that might be far from the respective surface (terrain profile or vegetation cover) but usually between those two surfaces. Such outliers correspond typical mismeasured points in LiDAR data.

The basic tests and numerical examples of LiDAR data processing confirm that AMS estimation should be considered an alternative for classical methods, especially when we suspect that observation sets are affected by outliers. TLS or ALS data often contain some mismeasured points; hence it is no doubt that such data might be processed by applying AMS method successfully.

Author contributions

Conceptualization: R.D. and P.W.; data curation: P.W.; formal analysis: R.D.; investigation: R.D. and P.W.; methodology: R.D. and P.W.; software: P.W.; supervision: R.D.; visualization: P.W.; writing – original draft: R.D. and P.W.; writing – review & editing: R.D. and P.W.

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This study was supported by the project "Innovative precise monitoring system based on integration of low-cost GNSS and IMU MEMS sensors" POIR 01.01.01-00-0753/21, co-financed by the European Regional Development Fund within the Sub-measure 1.1.1 of the Smart Growth Operational Program 2014-2020.

References

Amiri-Simkooei, A. (2003). Formulation of L1 norm minimization in Gauss-Markov models. *J. Surv. Eng.*, 129, 37–43. DOI: 10.1061/(ASCE)0733-9453(2003)129:1(37).

Baarda, W. (1968). *A testing procedure for use in geodetic networks*. In Publications on Geodesy 9, 5, pp. 97. Delft: Nethelands Geodetic Commision.

Baselga, S. (2007). Global optimization solution of robust estimation. *J. Surv. Eng.*, 133, 123–128. DOI: 10.1061/(ASCE)0733-9453(2007)133:3(123).

Baselga, S., and García-Asenjo, L. (2008). GNSS differential positioning by robust estimation. *J. Surv. Eng.*, 134, 21–25. DOI: 10.1061/(ASCE)0733-9453(2008)134:1(21).

Cabaleiro, M., Riveiro, B., Arias, P. et al. (2015). Algorithm for beam deformation modeling from LiDAR data. *Measurement*, 76, 20–31. DOI: 10.1016/j.measurement.2015.08.023.

Carrilho, A.C., and Galo, M. (2019). Automatic object extraction from high resolution aerial imagery with simple linear iterative clustering and convolutional neural networks. ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W16, 61–66.

Carrilho, A.C., Galo, M., and Santos, R.C. (2018). Statistical outlier detection method for airborne LiDAR data. ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1, 87–92.

Chang, X.-W., and Guo, Y. (2005). Huber's M-estimation in relative GPS positioning: computational aspects. *J. Geod.*, 79, 351–362. DOI: 10.1007/s00190-005-0473-y.

Crespo-Peremarch, P., Tompalski, P., Coops, N.C. et al. (2018). Characterizing understory vegetation in Mediterranean forests using full-waveform airborne laser scanning data. *Remote Sens. Environ.*, 217, 400–413. DOI: 10.1016/j.rse.2018.08.033.

Duchnowski, R. (2013). Hodges-Lehmann estimates in deformation analyses. *J. Geod.*, 87, 873–884. DOI: 10.1007/s00190-013-0651-2.

Duchnowski, R., and Wisniewski, Z. (2019). Robustness of Msplit(q) estimation: A theoretical approach. *Studia Geophys. et Geod.*, 63, 390–417. DOI: 10.1007/s11200-018-0548-x.

Duchnowski, R., and Wisniewski, Z. (2020). Robustness of squared Msplit(q) estimation: Empirical analyses. *Studia Geophys. et Geod.*, 64, 153–171. DOI: 10.1007/s11200-019-0356-y.

Duchnowski, R., and Wyszkowska, P. (2022). Empirical influence functions and their non-standard applications. *J. Appl. Geod.*, 16, 9–23. DOI: 10.1515/jag-2021-0012.

Ge, Y., Yuan, Y., and Jia, N. (2013). More efficient methods among commonly used robust estimation methods for GPS coordinate transformation. *Surv. Rev.*, 45, 229–234. DOI: 10.1179/1752270612Y.0000000028.

Gui, Q., and Zhang, J. (1998). Robust biased estimation and its applications in geodetic adjustments. *J. Geod.*, 72, 430–435. DOI: 10.1007/s001900050182.

Hekimoglu, S., and Berber, M. (2003). Effectiveness of robust methods in heterogeneous linear models. *J. Geod.*, 76, 706–713. DOI: 10.1007/s00190-002-0289-y.

Hekimoglu, S., Erdogan, B., and Erenoglu, R.C. (2015). A new outlier detection method considering outliers as model errors. *Exp. Tech.*, 39, 57–68. DOI: 10.1111/j.1747-1567.2012.00876.x.

Kargoll, B. (2005). *Comparison of some robust parameter estimation techniques for outlier analysis applied to simulated GOCE mission data*. In: C. Jekeli, L. Bastos, J. Fernandes (Eds.) International Association of Geodesy Symposia, Gravity, Geoid and Space Missions (pp. 77–82). Berlin: Springer.

Lehmann, R. (2013a). 3σ-rule for outlier detection from the viewpoint of geodetic adjustment. *J. Surv. Eng.*, 139, 157–165. DOI: 10.1061/(ASCE)SU.1943-5428.0000112.

Lehmann, R. (2013b). On the formulation of the alternative hypothesis for geodetic outlier detection. *J. Geod.*, 87, 373–386. DOI: 10.1007/s00190-012-0607-y.

Li, J., Wang, A., and Xinyuan, W. (2013). Msplit estimate the relationship between LS and its application in gross error detection. *Mine Surv.*, 2, 57–59.

Pope, A.J. (1976). *The statistics of residuals and the detection of outliers*. In.: NOAA Technical Report, NOS 65 (pp. 466). Rockville: U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Survey, National Geodetic Survey, Geodetic Research and Development Laboratory.

Wisniewski, Z. (2009). Estimation of parameters in a split functional model of geodetic observations (Msplit estimation). *J. Geod.*, 83, 105–120. DOI: 10.1007/s00190-008-0241-x.

Wisniewski, Z. (2010). Msplit(q) estimation: estimation of parameters in a multi split functional model of geodetic observations. *J. Geod.*, 84, 355–372. DOI: 10.1007/s00190-010-0373-7.

Wyszkowska, P., and Duchnowski, R. (2019). Msplit estimation based on L1 norm condition. *J. Surv. Eng.*, 145, 04019006. DOI: 10.1061/(ASCE)SU.1943-5428.0000286.

Wyszkowska, P., and Duchnowski, R. (2020a). Iterative process of Msplit(q) estimation. *J. Surv. Eng.*, 146, 06020002. DOI: 10.1061/(ASCE)SU.1943-5428.0000318.

Wyszkowska, P., and Duchnowski, R. (2020b). Systematic bias of selected estimates applied in vertical displacement analysis. *J. Geod. Sci.*, 10, 41–47. DOI: 10.1515/jogs-2020-0103.

Wyszkowska, P. (2021). *Uodporniona Msplit estymacja i jej podstawowe zastosowania (Robustified Msplit estimation and its basic applications)*. PhD dissertation (pp. 257). Olsztyn: University of Warmia and Mazury in Olsztyn.

Wyszkowska, P., Duchnowski, R., and Dumalski, A. (2021). Determination of terrain profile from TLS data by applying Msplit estimation. *Remote Sens.*, 13, 31. DOI: 10.3390/rs13010031.

Xu, P. (1989). On robust estimation with correlated observations. *Bull. Géodésique*, 63, 237–252. DOI: 10.1007/BF02520474.

Yang, Y. (1994). Robust estimation for dependent observations. *Manuscripta Geodaetica*, 19, 10–17.

Zienkiewicz, M.H. (2015). Application of Msplit estimation to determine control points displacements in networks with unstable reference system. *Surv. Rev.*, 47, 174–180. DOI: 10.1179/1752270614Y.0000000105.