Research Article

Global Optimization of Redescending Robust Estimators

Sergio Baselga [iD],[1] **Ivandro Klein** [iD],[2,3] **Stefano Sampaio Suraci** [iD],[4]
Leonardo Castro de Oliveira [iD],[4] **Marcelo Tomio Matsuoka** [iD],[5]
and Vinicius Francisco Rofatto [iD][5]

[1]*Department of Cartographic Engineering, Geodesy and Photogrammetry, Universitat Politècnica de València,*
 Camino de Vera s/n, 46022 València, Spain
[2]*Department of Civil Construction, Federal Institute of Santa Catarina, Florianopolis 88020-300, SC, Brazil*
[3]*Graduate Program in Geodetic Sciences, Federal University of Paraná, Curitiba 81531-990, PR, Brazil*
[4]*Cartographic Engineering Department, Military Institute of Engineering, Rio de Janeiro, Brazil*
[5]*Institute of Geography, Federal University of Uberlandia, Monte Carmelo 38500-000, MG, Brazil*

Correspondence should be addressed to Sergio Baselga; serbamo@cgf.upv.es

Robust estimation has proved to be a valuable alternative to the least squares estimator for the cases where the dataset is contaminated with outliers. Many robust estimators have been designed to be minimally affected by the outlying observations and produce a good fit for the majority of the data. Among them, the redescending estimators have demonstrated the best estimation capabilities. It is little known, however, that the success of a robust estimation method depends not only on the robust estimator used but also on the way the estimator is computed. In the present paper, we show that for complicated cases, the predominant method of computing the robust estimator by means of an iteratively reweighted least squares scheme may result in a local optimum of significantly lower quality than the global optimum attainable by means of a global optimization method. Further, the sequential use of the proposed global robust estimation proves to successfully solve the problem of M-split estimation, that is, the determination of parameters of different functional models implicit in the data.

1. Introduction

The statistical methods used for the adjustment of observations rely on several assumptions. Most commonly, the observed data are assumed to follow a normal distribution (also called Gaussian distribution) and the functional model relating the observed data with the unknown system parameters is assumed to hold exactly. Under these assumptions, it can be proved that the least squares (LS) estimator is the best linear unbiased estimator (Gauss–Markov theorem [1, 2]).

It is not uncommon, however, that the data collected contain one or several atypical observations, called outliers. They deviate from the general data pattern for some reason: they may be gross errors (observation values incorrectly noted down, for example) or pieces of data belonging to distributions other than the one of the main bulk of data (for example, due to the occasional unexpected observation of a different phenomenon concurrent with the main outcome of the experiment designed). It can also be that the functional model used to encapsulate the relationship between the unknown parameters and the observables is incomplete or simply approximate. In these cases, namely, the appearance of outliers or the existence of systematic effects in the observations or the functional model, the LS estimator may be very adversely influenced and produce a clearly unacceptable solution for the majority of the data.

While procedures to eliminate gross errors are routinely used (recent examples in geosciences include [3–5]); unfortunately, on certain occasions, some undesired errors may remain undetected and slip into the adjustment phase. Robust estimation techniques were devised in the last decades to attain a solution for the adjustment problem that is minimally

affected by the appearance of outliers in the data or systematic effects either in the data or in the model (inexactness of the functional model) while producing a solution close to the LS solution in the case of correct functional model and inexistence of outliers and systematic errors in the data. The reader may consult [6] for a general reference.

Many different robust estimators have been proposed over the last decades, some showing better performance than others in coping with undesired errors of different sizes and distributions. Little attention has gained, however, the question of how to compute the solution once the robust estimator has been selected, so that the Iteratively Reweighted Least Squares (IRLS) scheme is almost universally indicated as the procedure to follow (e.g., [6–8]). This may be due to the simplicity in recasting the new estimation techniques in an iterative procedure based on the familiar LS scheme; but, as we will see, this decision is at the cost of critically undermining the robust estimator capabilities, which are essential to successfully cope with the most complicated cases.

After a quick review of the types and properties of robust estimators, which will be needed to develop the subsequent sections, we will summarize the IRLS method to compute a robust estimator and propose a solution for this computation problem based on a Global Optimization algorithm (GO). We discuss then that a GO solution of the robust estimator is always equal or better than the solution by IRLS and can be referred to as global robust estimation, in contrast with the local robust estimation that is achieved by IRLS. The differences will be evident in the application section, whereas a novel result shows that global robust estimation efficiently solves the problem of M-split estimation resolved so far in a completely different way (see, for instance, [9]).

2. Types and Properties of Robust Estimators

We deal now with the so-called maximum likelihood estimators, or M-estimators, that is, the type of estimators that maximize the corresponding likelihood function over the parameter space. As Huber [10] proposed, the estimation problem entails the minimization over the entire parameter space of a function f of the parameters x and the adjustment residuals v

$$\widehat{x} = \arg\min_x f(v, x) = \arg\min_x \sum_{i=1}^{m} \rho(v_i, x). \qquad (1)$$

The function f, which is desired to be minimized, can be called the objective function. It is obtained as the sum of certain function ρ computed for every residual. This ρ function needs to have some desirable properties for the solution to be robust against outliers and systematic errors. The derivative of this ρ function, sometimes simply referred to as influence function, can be denoted by $\psi(v_i)$

$$\psi(v_i) = \frac{\partial \rho(v_i)}{\partial v_i}. \qquad (2)$$

The computation of the robust estimation problem has been routinely done by means of an IRLS scheme: from early

seminal works, [11], and early computing software packages, MIT ROSEPACK [12], until the most recent proposals, e.g., [13–15], including the general references on robust estimation, e.g., [6]. According to the IRLS scheme, in every iteration, the residuals of the previous iteration are used to form weights that are introduced in a new LS adjustment to obtain a new solution. The functions to obtain the corresponding weights $\omega(v_i)$ to solve the robust estimation problem by means of IRLS are easily obtained as follows:

$$\omega(v_i) = \frac{\psi(v_i)}{v_i}. \qquad (3)$$

For the case of the LS estimator, objective ρ, influence ψ, and weight ω functions are as follows:

$$\rho(v_i) = \frac{v_i^2}{2}, \qquad (4)$$

$$\psi(v_i) = v_i, \qquad (5)$$

$$\omega(v_i) = 1. \qquad (6)$$

The least L_1-norm or least absolute deviation estimator is significantly more robust against outliers and systematic errors. Its corresponding functions are as follows:

$$\rho(v_i) = |v_i|, \qquad (7)$$

$$\psi(v_i) = \text{sgn}(v_i), \qquad (8)$$

$$\omega(v_i) = \frac{1}{|v_i|}. \qquad (9)$$

It may be worth mentioning that if the estimation problem consists of obtaining the best value of a single quantity that has been repeatedly measured, the LS estimator is the arithmetic mean, and the least L_1-norm estimator is the median.

In general, we can choose values for p other than 1 and 2 and use the corresponding L_p-norm estimator, whose objective ρ, influence ψ, and weight ω functions are as follows:

$$\rho(v_i) = \frac{|v_i|^p}{p}, \qquad (10)$$

$$\psi(v_i) = sign(v_i)|v_i|^{p-1}, \qquad (11)$$

$$\omega(v_i) = |v_i|^{p-2}. \qquad (12)$$

Noninteger exponents may be used, for example $p = 1.5$, and some have even proposed to compute the best p value adapted to the sample [16].

$$p = \frac{9}{\beta_2^2},$$

$$\qquad\qquad (13)$$

$$\beta_2^2 = \frac{dof \sum_{i=1}^{m} v_i^4}{\left(\sum_{i=1}^{m} v_i^2\right)^2},$$

for number of degrees of freedom *dof* and residuals v_i of an initial adjustment.

Special combinations with some of these norms can also be used to obtain robust results (see, e.g., [17, 18]).

The L_1-norm estimator (all L_p-norms in fact) and many others that have been proposed and used in a relatively extensive way (e.g., Huber's estimator) lack, however, a desirable property that may be key to success in complicated cases: they are not redescending estimators.

The theory of redescending estimators originated in the work by Hampel et al. [19]. These are estimators having ψ functions that are not only zero in origin (the correct solution) but decrease toward zero far from the origin. They offer an increase in robustness toward large outliers and have a high breakdown point (in simple words, they resist the appearance of a very large proportion of outliers in the sample).

Some redescending estimators widely used are as follows:

(i) Geman–McClure estimator [20].

$$\rho(v_i) = \frac{v_i^2/2}{1 + v_i^2}, \tag{14}$$

$$\psi(v_i) = \frac{v_i}{\left(1 + v_i^2\right)^2}, \tag{15}$$

$$\omega(v_i) = \frac{1}{\left(1 + v_i^2\right)^2}. \tag{16}$$

(ii) Welsch estimator [21].

$$\rho(v_i) = \frac{c^2}{2}\left[1 - e^{-\left(v_i/c\right)^2}\right], \tag{17}$$

$$\psi(v_i) = v_i e^{-\left(v_i/c\right)^2}, \tag{18}$$

$$\omega(v_i) = e^{-\left(v_i/c\right)^2}, \tag{19}$$

where the constant c normally takes a value around 1 to 5.

(iii) Tukey's biweight estimator [22].

$$\rho(v_i) = \begin{cases} \dfrac{c^2}{6}\left(1 - \left[1 - (v_i/c)^2\right]^3\right) & \text{if } |v_i| \leq c \\[2ex] \dfrac{c^2}{6} & \text{if } |v_i| > c \end{cases}, \tag{20}$$

$$\psi(v_i) = \begin{cases} v_i\left[1 - (v_i/c)^2\right]^2 & \text{if } |v_i| \leq c \\ 0 & \text{if } |v_i| > c \end{cases}, \tag{21}$$

$$\omega(v_i) = \begin{cases} \left[1 - (v_i/c)^2\right]^2 & \text{if } |v_i| \leq c \\ 0 & \text{if } |v_i| > c \end{cases}, \tag{22}$$

where the threshold constant c normally takes a value around 1 to 5. Note that in the following section, we propose an equivalent redefinition of these functions that formally avoids piecewise definitions.

Figure 1 shows the ψ functions of all the above-mentioned robust estimators. As it can be seen, only Geman–McClure, Welsch, and Tukey are redescending estimators.

In the last years, redescending estimators have been extensively proved to outperform nonredescending estimators in the most complicated cases (see, e.g., [23]) but, as they are solved by IRLS, the initial solution used « must be robust in order to ensure convergence to a "good" solution» ([6], p. 40). Otherwise, the solution achieved is suboptimal. That is, it is not the best possible solution (global optimum) but the best only in the vicinity of the initial solution used for the IRLS. How to start with a good enough initial solution is still a case of research as it can be exemplified in the recent work [24]. The underlying problem, however, is indeed the use of IRLS, an iterative process strongly relying on the goodness of the starting solution due to the implicit dependence of the procedure on the least squares estimator.

We propose to abandon altogether the solution by IRLS in favor of a GO method, which eliminates the requirement of having a "good enough" initial solution. While the differences in computing the robust estimator by one procedure or the other (IRLS or GO) may be negligible in relatively easy problems (e.g., the case of levelling networks, [25]), in really complicated problems, one can definitely see a significant difference in favor of GO, which provides the global optimum and not simply a local optimum close to the initial solution, as in IRLS.

Global robust estimation, robust estimation where the estimator is computed by a GO method guaranteeing that the global optimum is attained, succeeds where the use of IRLS (local robust estimation) fails, as it will be shown. This applies to complicated examples in terms of large contamination by outliers and includes the possibility of successfully performing M-split estimation with no other tool than global robust estimation.

3. Methods

Let us solve the overdetermined system of observation equations

$$Ax = k + v, \tag{23}$$

where x is the vector of unknown parameters, A a coefficient matrix, k the vector of independent terms, and v the vector of adjustment residuals. The system of equations is assumed to have been previously reduced to the unit weight.
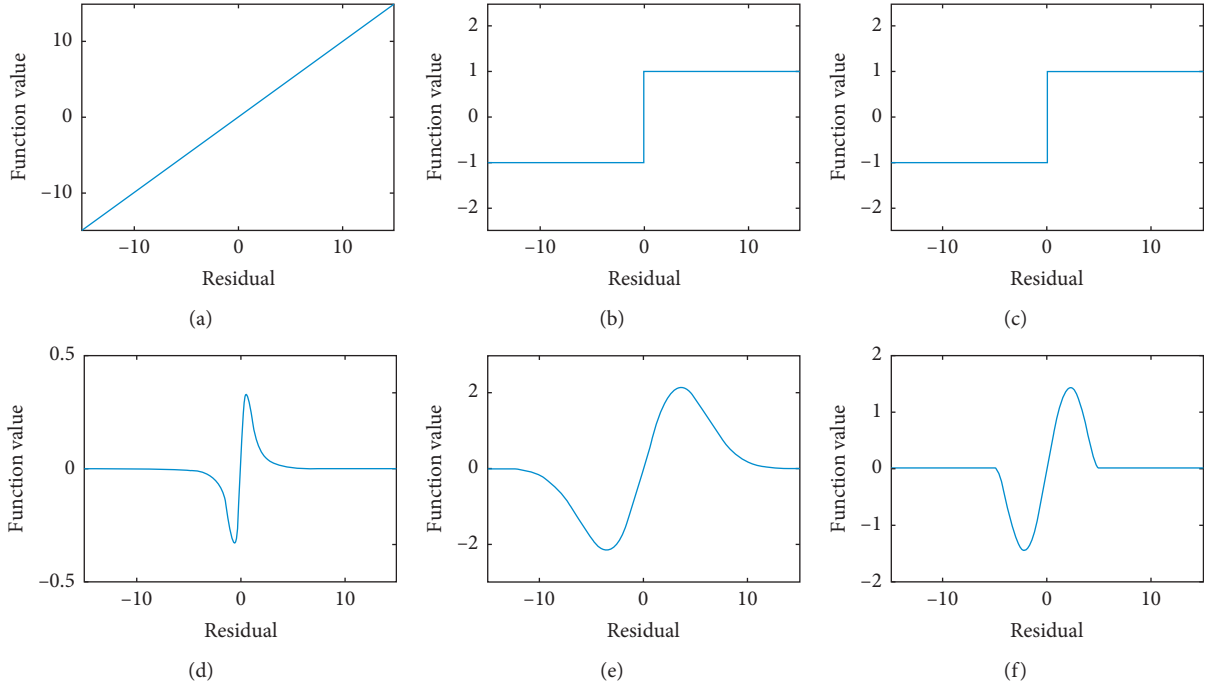
(a)

(b)

(c)

(d)

(e)

(f)

FIGURE 1: ψ functions of several robust estimators. (a) Least square estimator, (b) least L1-norm estimator, (c) least L1.5-norm estimator, (d) Geman–McClure estimators, (e) Welsch estimators ($c = 5$), and (f) Tukey estimators ($c = 5$).

3.1. Robust Estimation by IRLS.
The LS solution for the system of equations above, equation (23), is as follows:

$$x = \left(A^T A\right)^{-1} A^T k, \tag{24}$$

and the adjustment residuals can be computed by introducing the x vector obtained into

$$v = Ax - k. \tag{25}$$

These residuals can be used to obtain the weights for computing every robust estimator as an IRLS problem, using the corresponding previous equations (9), (12), (16), (19), or (22). The new solution for the system of weighted equations (having included the W weight matrix with weights ω_i in the diagonal) is as follows:

$$x = \left(A^T W A\right)^{-1} A^T W k, \tag{26}$$

and the new residuals

$$v = Ax - k, \tag{27}$$

which are used to compute new weights and new adjustment solutions in an iterative procedure that is stopped when no significant difference between iterations is found.

3.2. Robust Estimation by GO.
Alternatively to the use of IRLS for computing the robust estimator, we propose the use of a GO method, which, as shown in the next section, will produce a global optimum, sometimes significantly different from the mere local optimum achievable by IRLS for complicated estimation problems.

The GO method searches in a prescribed orderly way for the x vector minimizing the value of the objective function equation (1) using the corresponding expression ρ for the estimator (note this is also applicable to the LS estimator), say equations (4), (7), (10), (14), (17), or (20).

In the robust estimator function definitions above, note that Tukey's biweight estimator is piecewise-defined in terms of the residual achieved in the previous iteration. We want a unique expression and desire not to call values from previous iterations, so we propose the redefinition of equation (20) as follows:

$$\rho(v_i) = 1 - \left[1 - \left(\frac{v_i}{c}\right)^2\right]^3 H(c - v_i) H(c + v_i), \tag{28}$$

where we have omitted the constant factor $c^2/6$ for simplicity, since it has no influence on the final result, and we have introduced the Heaviside step function $H(x)$, which returns 0 for $x < 0$ and 1 for $x > 0$, and is customarily included in standard computation software (e.g., Matlab). With this definition, if either $v_i > c$ (then $c - v_i < 0$ and $H(c - v_i) = 0$) or $v_i < -c$ (then $c + v_i < 0$ and $H(c + v_i) = 0$), then the second summand vanishes as required.

Now, we describe an algorithm for computing any robust estimator by using the Simulated Annealing (SA) method, a GO method that has recently found many successful applications in the field of geosciences, e.g., [25–29]. The method is based on the analogy with the process of self-construction of crystalline networks. The interested reader should consult more detailed references (e.g., [30, 31]); here, we only show a particular adaptation to solve the problem at hand.

Step 1. A search domain should be defined and a point in it be taken as starting point. To this end, we can make use, for example (although this is not important), of the initial LS solution of the problem x_0 as the starting solution and define $x_0 \pm k_\sigma$ as the search domain

$$x \in [x_0 - k\sigma, x_0 + k\sigma], \qquad (29)$$

with k being an arbitrarily large value that guarantees (with a large degree of certainty) that the global optimum is inside the search domain and σ a vector with the standard deviations of every parameter, which are obtained as the square root of the values in the diagonal of the covariance matrix C_x

$$C_x = \frac{\nu^T \nu}{\text{dof}} (A^T A)^{-1}, \qquad (30)$$

where dof denotes the number of degrees of freedom in the adjustment (equal to the number of equations, m, minus the number of unknowns, n).

It is essential to note that unlike the case of IRLS, where the initial solution causes the final solution to be the closest local optimum, in the SA method, the starting point may be completely arbitrary since the final solution obtained will be independent of the initial point used. Also, we can take $k = 20$ or the larger the value we may desire at the only risk of increasing the computational burden. In other words, the use of the initial LS solution (or any other sensible alternative) to define both the starting value and the search domain is completely uncritical.

An initial movement amplitude, σ_0, needs to be defined. It can be taken as a fraction of the search space half-width mentioned above so that it enables to comfortably explore the entire search space in a few iterations, for example,

$$\sigma_0 = \frac{k}{5} \sigma. \qquad (31)$$

This initial movement amplitude σ_0 is taken as the current movement amplitude σ_{current}.

Step 2. The initial solution x_0 is taken both as the current solution x_{current} and the best solution x_{best}. Using these values, we compute the corresponding adjustment residuals ν, equation (25), and the corresponding value for the objective function

$$f(\nu) = \sum_{i=1}^{m} \rho(\nu_i). \qquad (32)$$

To this end, we make use of the desired form of the ρ function: equation (7) if the LS estimator is desired, equation (10) if the minimum L_1-norm estimator is desired, etc. The value initially obtained is stored both as the best (f_{best}) and the current (f_{current}) value of the f function.

Step 3. The current solution vector and the current objective function value are stored as the corresponding previous values: $x_{\text{current}} \longrightarrow x_{\text{previous}}$, $f_{\text{current}} \longrightarrow f_{\text{previous}}$

A new tentative solution is generated from the previous solution x_{previous} and the current movement amplitude σ_{current} as taken from the Normal distribution

$$x_1 \sim N(x_{\text{previous}}, \sigma_{\text{current}}). \qquad (33)$$

We can do this computation element by element that is, for the first element of vector x_1, we draw a value from a normal distribution of mean, the first element of x_{previous}, and standard deviation, the first element of σ_{current}, and analogously for the rest of the elements.

If the solution x_1 obtained is located inside the search domain, equation (29), it is a feasible solution. Otherwise, it is discarded and a new solution is drawn from the Normal distribution, equation (33).

The feasible solution may be accepted or not as the current solution for subsequent iterations, depending on certain conditions (akin to the process observed in nature by which crystalline networks are self-constructed) which are studied in the following step.

Step 4. We compute the residuals ν corresponding to solution x_1 and evaluate therein the objective function f analogously as done before in Step 2. This becomes the current value for f function, f_{current}.

  (i) If the solution is better than the previous one, that is, if $f_{\text{current}} < f_{\text{previous}}$, the solution is accepted, that is, x_1 is stored as x_{current}: $x_1 \longrightarrow x_{\text{current}}$. It must also be checked whether the current value of the objective function is the best value so far and rename variables if this is the case ($f_{\text{current}} \longrightarrow f_{\text{best}}$ and $x_{\text{current}} \longrightarrow x_{\text{best}}$)

  (ii) If the solution is not better than the previous one, that is, if $f_{\text{current}} \geq f_{\text{previous}}$, the solution is only accepted with a very low probability p as to emulate the very rare cases observed in nature where worse energy states are occasionally accepted in the process of crystalline network self-construction. In short, if a number drawn at random from the uniform distribution [0,1] is below, say, 0.01 (which only happens with 1% of probability); then, $x_1 \longrightarrow x_{\text{current}}$; otherwise, x_{current} is taken from the previous iteration (also $f_{\text{previous}} \longrightarrow f_{\text{current}}$).

Step 5. Cooling process. The amplitude of movement is reduced as emulating the cooling process during which the crystalline network is created. The cooling equation can be as simple as follows:

$$\sigma_{\text{current}} = \beta^{iterNo} \sigma_0, \qquad (34)$$

where σ_0 is the initial movement amplitude, *iterNo* the iteration number, and β a number sufficiently close to unity (e.g., $\beta = 0.99999$), taking into account the fact that the closer to the unity, the slower the cooling and the corresponding computational process.

If the current movement amplitude can be considered negligible for the problem at hand, the algorithm stops. Otherwise, we return to Step 3.

It can be demonstrated that the SA method converges to the global optimum in probabilistic terms [30], and for this, we need a sufficiently slow cooling scheme (β sufficiently close to one). Although there can be no complete certainty that the solution attained is indeed the global optimum, a pair of hints reinforce the credibility of the solution: first, the final current solution must be coincident with the best solution attained all along the iterations; second, and important to check, subsequent executions of the algorithm (which use different numbers due to the intrinsic random nature of the algorithm) may produce the same result: the expected global optimum. In addition, we can also execute the algorithm with different initial solutions than the LS solution to verify that it also converges to the same result.

4. Application

4.1. Robust Estimation in the presence of Outliers. As said before, the computation of the robust estimator as an IRLS problem is the procedure customary found in the literature, also for application to geodetic problems: examples of different types include [32–34].

Depending on the problem complexity, the solution obtained by IRLS may already be the global optimum (thus coincident with the one obtained by GO). This is the case when the initial solution to start the IRLS process—the LS solution—is not extremely affected by the existing undesired errors and is relatively close to the correct solution. As shown in [25], the case of levelling networks appears to be one of such problems where robust estimation in the presence of outliers is equally well solved by IRLS and by GO since both are chiefly limited by the relatively low redundancy of the problem.

We turn our attention now to a problem where IRLS and GO behave very differently. It is a linear adjustment taken from Maronna et al. ([6], p.147 Table 5.6) based, in its turn, on a work by Jalali-Heravi and Konouz [35]. The experiment consists of the observation of different chemical compounds of two quantities (heat and Krafft point), which are in theory linearly related, where a significant proportion of outlying observations exists. In fact, some of the outliers correspond to a different structure than the principal one, whereas others may be gross errors.

Figure 2 and Tables 1 and 2 show the results obtained by the estimators presented before (those whose ψ functions were depicted in Figure 1). For GO, they were obtained by the SA method with the following parameters: cooldown factor $\beta = 0.99999$, low probability $p = 0.01$, and for each unknown: search space half-width $k\sigma$ with σ from the LS estimation and $k = 300$, initial movement amplitude $\sigma_0 = k\sigma/5$ (with the same values for σ and k), and final movement amplitude $\sigma_0/10000$. The following results were obtained with the same values for all the decimal places shown after several runs (indeed with negligible discrepancies of the order of the tuneable final movement amplitude only). For the starting solution, we used the LS estimation values, although, as expected, it was checked in several test runs that the starting values do not affect the final results either.

The LS estimator computed by IRLS is simply the standard computation of LS since the equivalent weights are 1, recall equation (6). This result completely coincides with the computation of LS by GO, as it can be seen in the corresponding columns of Table 1. This is also the reason why only a solid line can be seen in the LS subplot in Figure 1 (the discontinuous line is superimposed on the solid one). As expected, the LS estimator is strongly affected by the existing outliers and is unable to make a good fit for the main bulk of the data.

The situation is quite similar for the minimum L_1 and minimum $L_{1.5}$ norms: in this problem, they are unable to provide a satisfactory solution for the majority of the data, and they produce a compromise solution. In both cases, the IRLS and the GO methods yield the same results despite minimum, negligible discrepancies. LS, minimum L_1, and minimum $L_{1.5}$ norms produce relatively similar solutions among them, all incapable of distinguishing the main bulk of the data from their outlying observations.

The situation changes with the Geman–McClure estimator. We can see, first, that the solution is completely different if computed by IRLS (discontinuous line in Figure 2) or by GO (solid line). The Geman–McClure estimator computed by IRLS is not really different from LS or the previous estimators: it fails to fit the line to the majority of the data. The Geman–McClure estimator computed by GO, however, performs more sensitively to the main linear substructure. Looking at the corresponding best values of the Geman–McClure objective function f obtained by IRLS and GO, 14.245 and 13.009, respectively, it is clear that the IRLS method has produced a suboptimal result, that is only a local optimum in the neighborhood of its starting solution, whereas the GO method has produced a better result (presumably the global optimum).

In the case of the Welsch estimator, we also see that the solution is poor if computed by IRLS (only a local optimum of the objective function f: 23.755), whereas it is very satisfactory if computed by GO (global optimum of the objective function f: 20.899).

This is analogous to the case of Tukey, where the robust estimator successfully captures the main linear substructure in the data but only if properly computed as GO, the IRLS scheme only producing a local optimum for Tukey's objective function.

In summary, we have experienced that, under a relatively high degree of contamination of the observations (by gross errors or minor substructures in the data), only the Geman–McClure (to a moderate extent), Welsch, and Tukey estimators succeed in finding a good fit to the majority of the data, and only as long as they are computed by GO (while they fail if computed by IRLS). Recall from Figure 1 that these are redescending estimators, having better estimation capabilities, as expected, but also strongly dependant on the starting solutions for their IRLS schemes.

Consequently, for cases where there is a large proportion of outliers with respect to the main bulk of the data, we propose not only the use of the best robust estimators, that is, redescending robust estimators like Welsch and Tukey, but also their computation in a GO scheme, inasmuch as their computation by means of IRLS may yield a local optimum only.
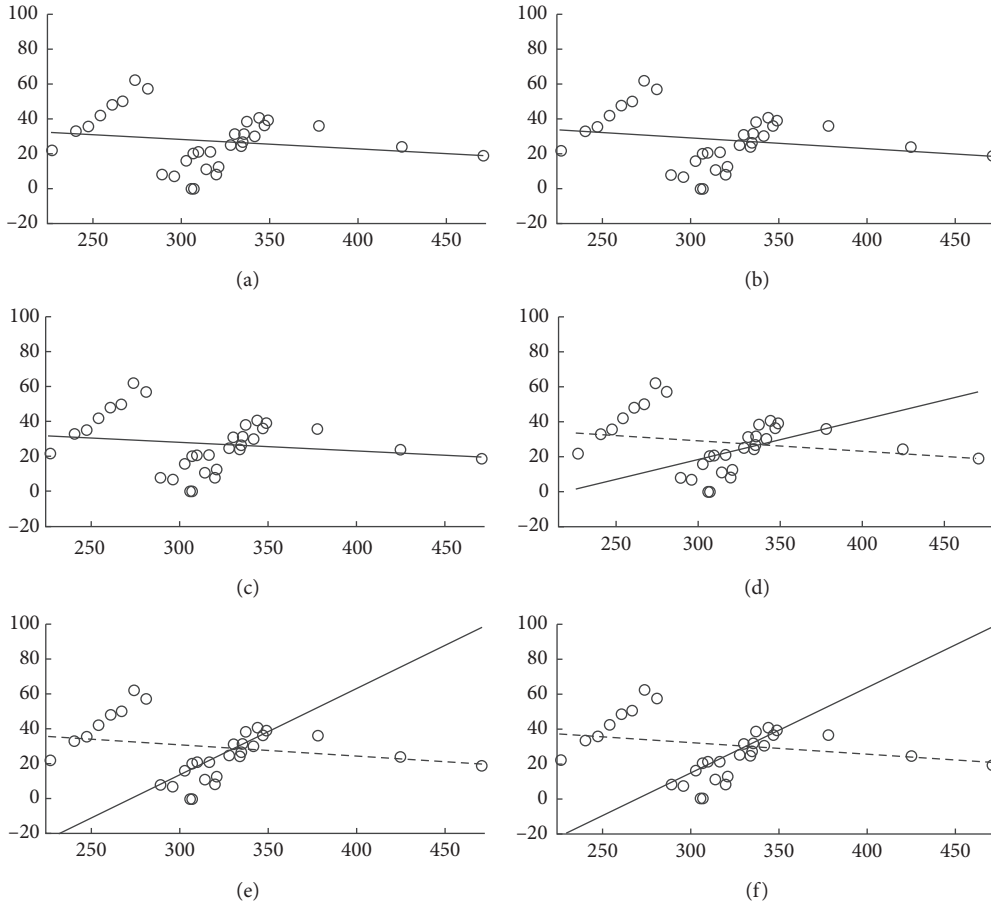
FIGURE 2: Lines fitted to the data by different estimators: the solid line is the result of the computation by GO and the discontinuous line by IRLS. (a) Least Squares, (b) Least L1 norm, (c) Least L1.5 norm, (d) Geman -McClure, (e) Welsch, (f) Tukey.

TABLE 1: Defining parameters (a and b) of lines $y = ax + b$ fitted to the data obtained with different nonredescending estimators computed by GO and by IRLS and corresponding best values of the objective function f.

	LS		L_1-norm		$L_{1.5}$-norm	
	IRLS	GO	IRLS	GO	IRLS	GO
a	−0.05626	−0.05626	−0.06074	−0.06074	−0.04952	−0.04952
b	45.1045	45.1045	47.6074	47.6074	43.2558	43.2555
f_{best}	3618.175	3618.175	388.227	388.227	1084.595	1084.595

TABLE 2: Defining parameters (a and b) of lines $y = ax + b$ fitted to the data obtained with different redescending estimators computed by GO and by IRLS and corresponding best values of the objective function f.

	Geman–McClure		Welsch		Tukey	
	IRLS	GO	IRLS	GO	IRLS	GO
a	−0.06036	0.22425	−0.06338	0.49127	−0.06766	0.48653
b	47.3326	−48.6879	49.8820	−133.7073	52.2803	−131.4851
f_{best}	14.245	13.009	23.755	20.899	26.353	24.530

4.2. Sequential Robust Estimation to Solve M-Split Estimation.

After having examined the previous example, one may wonder whether the main structure could be eliminated from the data after its successful identification and additional searches for minor substructures be performed then by the same procedure. This suggests the possible use of global robust estimation for what is known as M-split

estimation: the estimation of parameters of two or more unknown functional models from the observation of some existing mixed data (see, e.g., [9, 36]).

We study an example with possible interest in geomatics, especially in the processing of terrestrial laser scanning data: the adjustment of data to several circles whose number, origin, and radius are unknown. The idea, as well as the size

and location of the circles, has been taken from [37], but the data generated are different in our example since they include their particular random errors. The interested reader can find the data used in the Supplementary Information file.

These data correspond to the (x, y) coordinates in meters of 1050 points belonging to 7 circles with the same centers and radii as those in [37]. Here, the number of points computed for every circle has been selected so that it is proportional to the circumference length and, in order to make the problem a bit more complicated, the distribution of points in every circle is not equally spaced but taken with azimuth at random, as well as the corresponding radius being modified by a Gaussian noise of zero mean and 0.04 cm standard deviation. Figure 3 shows the generated data points and Table 3 the corresponding 7 circles.

All points (x_i, y_i) belonging to a circumference of center (x_c, y_c) and radius r satisfy the following equation:

$$(x_i - x_c)^2 + (y_i - y_c)^2 = r^2, \tag{35}$$

which can be transformed into the equivalent equation:

$$P_1 x_i + P_2 y_i + P_3 = -\left(x_i^2 + y_i^2\right), \tag{36}$$

with

$$P_1 = -2x_c, \tag{37}$$

$$P_2 = -2y_c, \tag{38}$$

$$P_3 = x_c^2 + y_c^2 - r^2, \tag{39}$$

where the circumference is now described by its parameters (P_1, P_2, P_3) instead of (x_c, y_c, r) with the advantage that the model equation (36) is linear in the parameters to estimate.

We perform the adjustment of the 1050 data points to the model equation (36) in order to obtain the parameters of the circle (or circles) by means of the same estimators as in the previous example and obtain the results displayed in Tables 4 and 5.

As with the previous example, the LS, minimum L_1 norm, and minimum $L_{1.5}$ norm estimators, computed both by IRLS and by GO, fail to produce a result that fairly represents any of the existing circles while discarding the rest of the observations. Only the redescending estimators (Geman–McClure, Welsch, and Tukey) produce a successful result as long as they are computed by GO (they also fail otherwise): they yield the parameters of circle No. 2, presumably because it is the existing substructure having more data points.

It is important to note that since the data points belonging to the circles were generated using the theoretical radius plus Gaussian noises of standard deviation 0.04 m, the parameters expected are not exactly those in Table 3 but others relatively close to them. We must therefore be satisfied with the P_1, P_2, P_3 values in Table 5 obtained by GO for Geman–McClure, Welsch, and Tukey estimators, which, in this case, correspond to circle No. 2. For the P_1, P_2, P_3 parameters obtained for circle No. 2, we can easily get the corresponding center (x_c, y_c) and radius r via equations
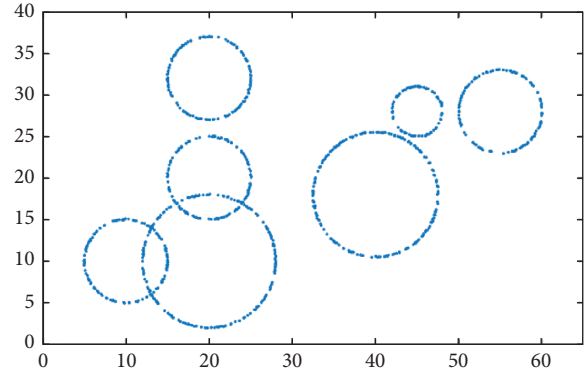

FIGURE 3: Generated data points.

(37)–(39). The comparison could also be made in terms of these x_c, y_c, r, as it will also be provided in subsequent tables.

Now we propose to discard all points in the dataset that can be ascribed to circle No. 2. It is worth noting that, due to the existence of a large number of redundant observations, we will prefer to delete a few points that possibly do not belong to circle No. 2 rather than to leave points of this circle that will not be able to be correctly ascribed to any of the remaining circles in subsequent computations.

To this end, for any point (x_i, y_i) in the dataset, we analyze the following condition derived from equation (35) for the values of center (x_c, y_c) and radius r previously obtained:

$$\left| \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} - r \right| < 10\sigma, \tag{40}$$

where the standard deviation σ is 0.04 m as before. If this condition is fulfilled, the point is assumed to belong to the circumference whose parameters have been obtained and is eliminated from the dataset to be analyzed in subsequent computations. Recall that, as explained above, we want to eliminate all points belonging to the first circle at the risk of eliminating others too, which would constitute no problem due to a large number of redundant observations. This is the reason for using such a high threshold as 10σ in equation (40). For lower values, like 3σ or 5σ, we observed that a very small percentage of the data points underwent the entire process with no ascription to any solution.

For all the remaining points in the dataset, the process is repeated. That is, a system of equations of the type of equation (36) is again formed and solved by the redescending estimators (Geman–McClure, Welsch, and Tukey) by means of GO. Obviously, the other estimators computed either by GO or IRLS or the redescending estimators computed by IRLS are no longer considered since they were already unable to yield the parameters of the first circle. This computation is iterated until the dataset is empty or has a very low number of points (say less than 0.1% of the original total number of points).

For the Geman–McClure estimator, the results obtained in the subsequent iterations are shown in Table 6, and the misfits, left-hand side of equation (40), for the corresponding datasets in each iteration are shown in Figure 4. As expected, in every iteration, the solution produces close-to-zero misfits for a group of observations, those of the corresponding circle.

TABLE 3: Circles and corresponding parameters: for each circle (circle no.), x_c and y_c are the coordinates of the center, r the radius, No. Points the number of points generated for the circle, and P_1, P_2, P_3 the circle parameters explained below.

Circle no.	x_c	y_c	r	No. points	P_1	P_2	P_3
1	10.0	10.0	5.0	136	−20.0	−20.0	175
2	20.0	10.0	8.0	219	−40.0	−20.0	436
3	20.0	20.0	5.0	136	−40.0	−40.0	775
4	20.0	32.0	5.0	136	−40.0	−64.0	1399
5	55.0	28.0	5.0	136	−110.0	−56.0	3784
6	45.0	28.0	3.0	82	−90.0	−56.0	2800
7	40.0	18.0	7.5	205	−80.0	−36.0	1868

All points (x_i, y_i) belonging to a circumference of center (x_c, y_c) and radius r satisfy the following equation:

TABLE 4: Circle parameters (P_1, P_2, and P_3) fitted to the data obtained with different nonredescending estimators computed by GO and by IRLS and corresponding best values of the objective function f.

	LS		L_1-norm		$L_{1.5}$-norm	
	IRLS	GO	IRLS	GO	IRLS	GO
P_1	−63.096	−63.096	−60.860	−60.862	−61.924	−61.924
P_2	−38.620	−38.620	−38.072	−38.047	−38.755	−38.755
P_3	1043.74	1043.74	1004.87	1004.53	1031.39	1031.38
f_{best}	25766225	25766225	185947	185947	2003060	2003060

TABLE 5: Circle parameters (P_1, P_2, and P_3) fitted to the data obtained with different redescending estimators computed by GO and by IRLS and corresponding best values of the objective function f.

	Geman–McClure		Welsch		Tukey	
	IRLS	GO	IRLS	GO	IRLS	GO
P_1	−63.136	−39.995	−62.921	−39.998	−62.882	−39.998
P_2	−38.527	−19.992	−38.626	−20.010	−38.632	−19.996
P_3	1043.73	435.83	1038.20	436.06	1036.38	435.92
f_{best}	522.05	436.80	1039.69	826.86	1041.64	836.92

TABLE 6: Circles and corresponding parameters obtained with the Geman–McClure estimator computed by GO: for each solution (Solution No.), x_c and y_c are the coordinates of the center, r the radius, L the circumference length, No. Points: the number of points generated for the circle, and P_1, P_2, P_3 are the circle parameters. Circle No. is the corresponding circle number in Table 3.

Solution no.	x_c	y_c	r	P_1	P_2	P_3	Circle no.
1	19.997	9.996	7.999	−39.995	−19.992	435.833	2
2	19.996	31.996	5.003	−39.992	−63.993	1398.593	4
3	39.971	18.025	7.489	−79.942	−36.049	1866.502	7
4	9.967	9.974	5.027	−19.934	−19.947	173.534	1
5	55.000	28.000	4.998	−110.001	−56.000	3784.077	5
6	20.038	20.003	4.987	−40.075	−40.006	776.755	3
7	44.994	27.993	3.004	−89.987	−55.987	2799.023	6

As it can be seen, the Geman–McClure estimator computed by GO produces along the sequential adjustments the circles No. 2, 4, 7, 1, 5, 3, and 6. Quite expectedly, they are obtained in the order of higher to lower number of corresponding points in the dataset.

The results are quite similar for the Welsch estimator by GO, Table 7 and Figure 5. Although the circles sequentially obtained are now circles No. 2, 7, 4, 1, 5, 3, and 6, they are still obtained from higher to the lower number of points in the dataset (circles 4, 7, 1, and 5 have the same number of points).

Finally, the results obtained for the Tukey estimator are shown in Table 8 and Figure 6. Although the parameters obtained for the circles are very similar to the ones

obtained by the previous two estimators, it is unknown why the order in which they are sequentially obtained is now circles No. 2, 1, 7, 6, 4, 3, and 5, which does not strictly correspond to the highest-to-lowest number of points in the dataset. This has little importance, however, inasmuch as the seven functional models which are implicit in the data have been revealed.

In summary, only the three redescending estimators (Geman–McClure, Welsch, and Tukey) computed by GO have succeeded in revealing the different functional models implicit in the data, that is, the seven circles, yielding their corresponding parameters within their expected accuracy of the order of a few cm.
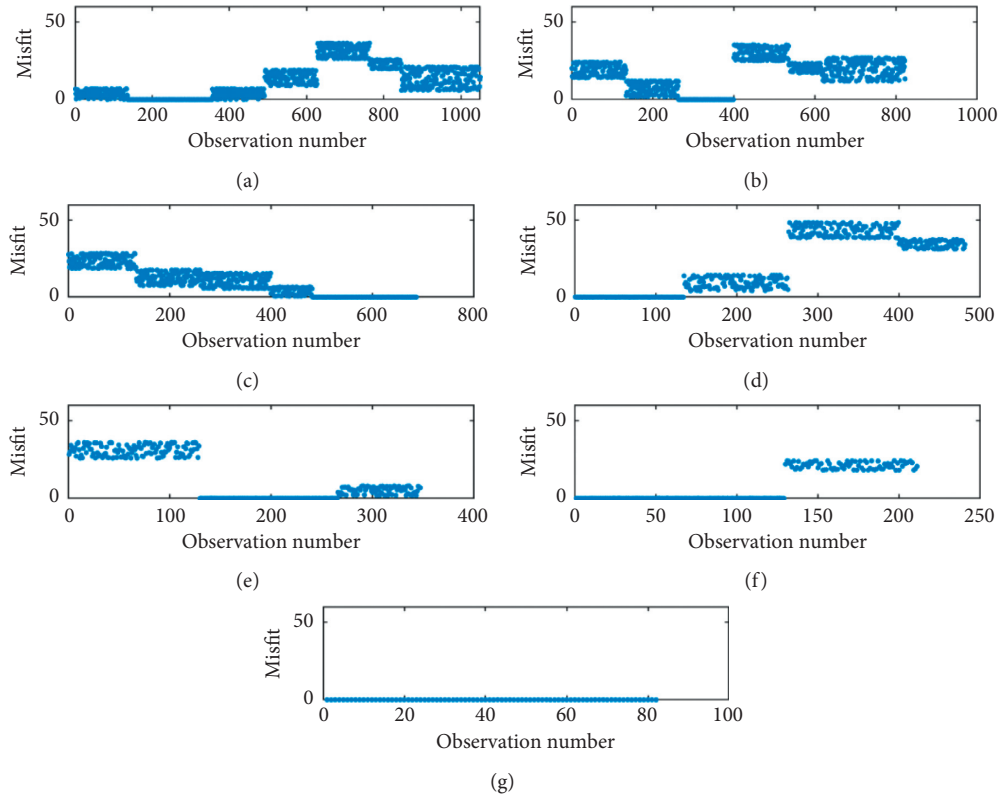
Figure 4: Misfits in the sequential adjustments with the Geman–McClure estimator computed by GO. (a) Solution 1, (b) Solution 2, (c) Solution 3, (d) Solution 4, (e) Solution 5, (f) Solution 6, and (g) Solution 7.

Table 7: Circles and corresponding parameters obtained with the Welsch estimator computed by GO: for each solution (Solution No.), x_c and y_c are the coordinates of the center, r the radius, L the circumference length, No. Points is the number of points generated for the circle, and P_1, P_2, P_3 are the circle parameters. Circle No. is the corresponding circle number in Table 3.

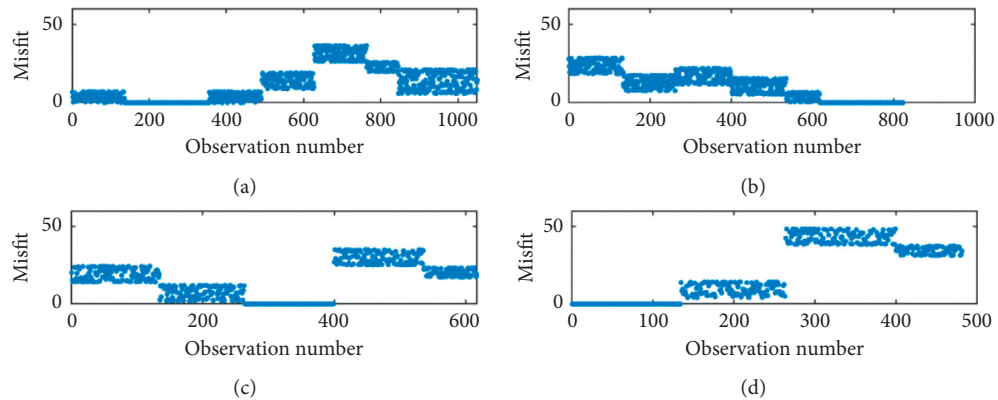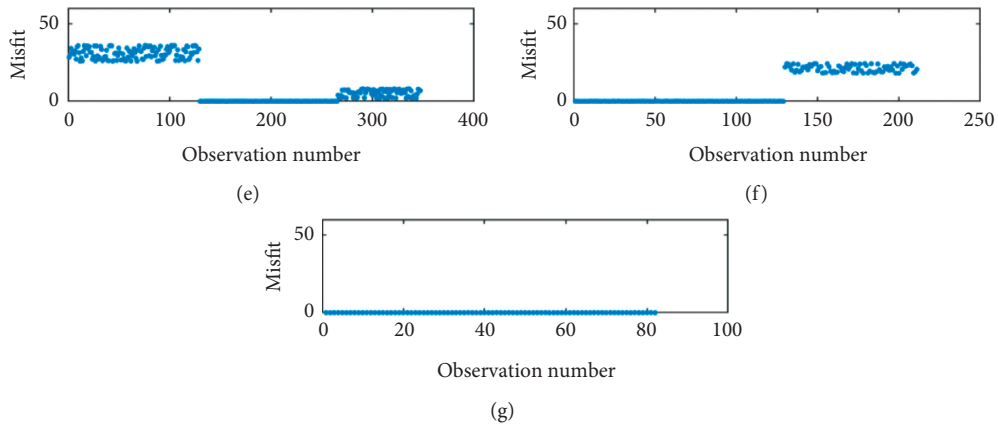| Solution no.	x_c	y_c	r	P_1	P_2	P_3	Circle no.
1	19.999	10.005	8.000	−39.998	−20.010	436.060	2
2	40.002	18.040	7.499	−80.004	−36.081	1869.377	7
3	20.063	32.008	4.973	−40.127	−64.015	1402.286	4
4	10.006	9.988	4.986	−20.012	−19.976	175.026	1
5	55.008	28.010	4.997	−110.016	−56.019	3785.429	5
6	20.000	20.001	4.998	−40.001	−40.002	775.079	3
7	44.994	27.996	3.004	−89.987	−55.993	2799.210	6

Figure 5: Continued.

(e)

(f)

(g)

FIGURE 5: Misfits in the sequential adjustments with the Welsch estimator computed by GO. (a) Solution 1, (b) Solution 2, (c) Solution 3, (d) Solution 4, (e) Solution 5, (f) Solution 6, and (g) Solution 7.

TABLE 8: Circles and corresponding parameters obtained with the Tukey estimator computed by GO: for each solution (Solution No.), x_c and y_c are the coordinates of the center, r the radius, L the circumference length, No. Points is the number of points generated for the circle, and P_1, P_2, P_3 are the circle parameters. Circle No. is the corresponding circle number in Table 3.

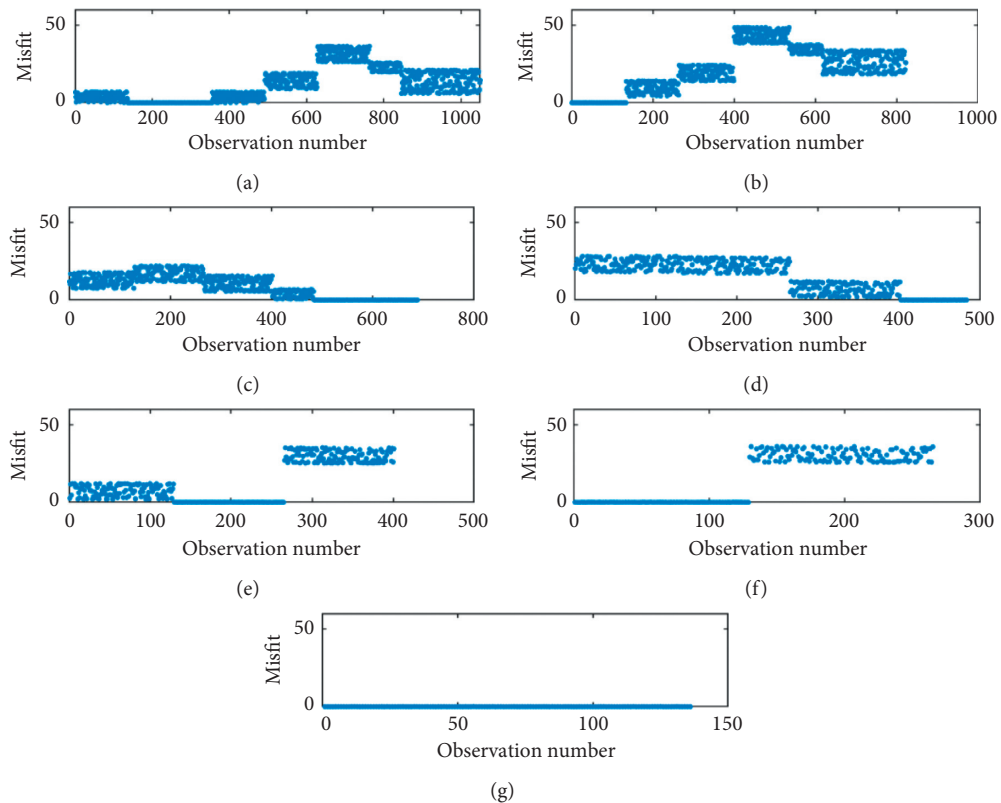| Solution no.	x_c	y_c	r	P_1	P_2	P_3	Circle no.
1	19.999	9.998	8.000	−39.998	−19.996	435.917	2
2	10.001	9.994	4.995	−20.003	−19.988	174.963	1
3	39.987	17.994	7.498	−79.974	−35.989	1866.540	7
4	44.970	27.980	3.018	−89.940	−55.960	2796.056	6
5	19.997	31.998	5.001	−39.994	−63.995	1398.704	4
6	20.000	20.000	4.999	−39.999	−40.000	774.998	3
7	55.003	28.003	4.997	−110.006	−56.005	3784.475	5

(a)

(b)

(c)

(d)

(e)

(f)

(g)

FIGURE 6: Misfits in the sequential adjustments with the Tukey estimator computed by GO. (a) Solution 1, (b) Solution 2, (c) Solution 3, (d) Solution 4, (e) Solution 5, (f) Solution 6, and (g) Solution 7.

5. Conclusions

In the present paper, we have shown that the success of a robust estimation method depends not only on the robust estimator used but also on the way the estimator is computed. We have shown that in problematic cases, the predominant method of computation by means of IRLS may result in a local optimum of poor quality.

This problem, which is a current topic of research, has been proposed to be solved in the literature by requiring a robust initial solution for the first iteration of the IRLS method. We have proposed here a different alternative: the computation of the robust estimator by a GO method. A particular implementation of the SA method has also been given for this purpose.

For cases where there is a large proportion of outliers with respect to the majority of the data, the use of the best robust estimators, redescending estimators like Geman–McClure, Welsch, and Tukey estimators, were proposed along with their computation in a GO scheme, inasmuch as their computation by means of IRLS yielded local optima only. We also propose an original, equivalent redefinition of some estimators, like Tukey's biweight estimator, in order to avoid the use of piecewise definitions for the sake of an easier implementation in a GO method.

The sequential use of these three redescending estimators—Geman–McClure, Welsch, and Tukey estimators—computed by GO was proved to successfully solve the M-split estimation, that is, the determination of parameters of different functional models implicit in the data. These estimators failed if computed by IRLS, in the same way, that other nonredescending robust estimators (least L_1 and least $L_{1.5}$ norms) as well as the LS estimator, irrespectively of the computation method, failed too when asked to determine any of the different functional models underlying the data sample.

In short, for an adjustment problem where the dataset is highly contaminated with outliers or there are several underlying functional models for the observations, we propose, as an original, novel alternative not to be found in the literature so far, the use of redescending estimators computed by GO as the best alternative to attain a good fit to the majority of the data. If desired, sequential adjustments may also reveal the corresponding parameters of the existing different functional models. As a future work, the effectiveness of this technique should be further analyzed in more challenging problems also with the application of other methods for GO apart from SA.

Data Availability

The data used for this study can be found in Maronna et al. ([6], p.147 Table 5.6) for Section 4.1 and the Supplementary Information file for Section 4.2.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Supplementary Materials

The x_y_7_circles.txt ASCII file contains the x and y coordinates of the points used for the example of 7 circles used in Section 4.2. (*Supplementary Materials*)

References

[1] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Upper Saddle River, NJ, USA, 2002.

[2] M. Hallin, "Gauss–Markov Theorem in Statistics," *Wiley StatsRef: Statistics Reference Online*, John Wiley and Sons, Ltd, Hoboken, NJ, USA, 2002.

[3] H. A. Abd-Elmotaal and N. Kühtreiber, "Suitable gravity interpolation technique for large data gaps in Africa," *Studia Geophysica et Geodaetica*, vol. 63, no. 3, pp. 418–435, 2019.

[4] W. Gao, J. Guo, M. Zhou, H. Yu, X. Chen, and B. Ji, "Gravity tides extracted from SSA-denoised superconducting gravity data with the harmonic analysis: a case study at Wuhan station, China," *Acta Geod Geophys*, vol. 55, 2020.

[5] V. F. Rofatto, M. T. Matsuoka, I. Klein, M. R. Veronez, M. L. Bonimani, and R. Lehmann, "A half-century of Baarda's concept of reliability: a review, new perspectives, and applications," *Survey Review*, vol. 52, no. 372, pp. 261–277, 2020.

[6] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, John Wiley and Sons, Chichester, UK, Wiley Series in Probability and Statistics, 2006.

[7] B. Kargoll, G. Kermarrec, J. Korte, and H. Alkhatib, "Self-tuning robust adjustment within multivariate regression time series models with vector-autoregressive random errors," *Journal of Geodesy*, vol. 94, no. 5, p. 51, 2020.

[8] L. Yang and Y. Shen, "Robust M estimation for 3D correlated vector observations based on modified bifactor weight reduction model," *Journal of Geodesy*, vol. 94, no. 3, p. 31, 2020.

[9] Z. Wisniewski, "$M_{\text{split}(q)}$ estimation: estimation of parameters in a multi split functional model of geodetic observations," *J Geod*, vol. 84, pp. 355–372, 2010.

[10] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[11] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics—Theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[12] V. Klema, "ROSEPACK-robust statistics estimation package," *ACM Signum Newsletter*, vol. 13, no. 2, pp. 18-19, 1978.

[13] L. Yao, P. Lin, J. Gao, and C. Liu, "Robust prediction algorithm based on a general EIV model for multiframe transformation," *Mathematical Problems in Engineering*, vol. 2019, Article ID 5173956, 10 pages, 2019.

[14] J. Li, Q. Hu, and M. Ai, "Robust geometric model estimation based on scaled Welsch q-norm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5908–5921, 2020.

[15] Y. Li, L. Hou, Y. Yang, and J. Tong, "Huber's M-estimation-based cubature kalman filter for an INS/DVL integrated system," *Mathematical Problems in Engineering*, vol. 2020, Article ID 1060672, 12 pages, 2020.

[16] B. S. Bor-Sen Chen, J. M. Jeng-Ming Chen, and S. C. Shen-Ching Shern, "An ARMA robust system identification using a generalized l/sub p/norm estimation algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1063–1073, 1994.

[17] Q. Liu, C. Yang, Y. Gu, and H. C. So, "Robust sparse recovery via weakly convex optimization in impulsive noise," *Signal Processing*, vol. 152, pp. 84–89, 2018.

[18] Q. Liu, Y. Gu, and H. C. So, "Smoothed sparse recovery via locally competitive algorithm and forward Euler discretization method," *Signal Processing*, vol. 157, pp. 97–102, 2019.

[19] F. R. Hampel, P. J. Rousseeuw, and E. Ronchetti, "The change-of-variance curve and optimal redescending M-estimators," *Journal of the American Statistical Association*, vol. 76, no. 375, pp. 643–648, 1981.

[20] S. Geman, D. E. McClure, and D. Geman, "A nonlinear filter for film restoration and other problems in image processing," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 4, pp. 281–289, 1992.

[21] J. E. Dennis and R. E. Welsch, "Techniques for nonlinear least squares and robust regression," *Commun Stat-Simul Comput*, vol. 7, pp. 345–359, 1976.

[22] A. E. Beaton and J. W. Tukey, "The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data," *Technometrics*, vol. 16, no. 2, pp. 147–185, 1974.

[23] G. Shevlyakov, J. W. Lee, K. M. Lee, V. Shin, and K. Kim, "Robust detection of a weak signal with redescending *M*-estimators: a comparative study," *Int. J Adapt Control Signal Process*, vol. 24, pp. 33–40, 2010.

[24] C. Agostinelli, M. Valdora, and V. J. Yohai, "Initial robust estimation in generalized linear models," *Computational Statistics and Data Analysis*, vol. 134, pp. 144–156, 2019.

[25] S. Baselga, I. Klein, S. S. Suraci, L. C. de Oliveira, M. T. Matsuoka, and V. F. Rofatto, "Performance comparison of least squares, iterative and global L1 norm minimization and exhaustive search methods for outlier detection in leveling networks," *Acta Geodynamica et Geomaterialia*, vol. 17, pp. 425–438, 2020.

[26] F. Jia and D. Lichti, "A comparison of simulated annealing, genetic algorithm and particle swarm optimization in optimal first-order design of indoor TLS networks," *ISPRS-ann Photogram Remote Sens Spat Inf Sci*, vol. IV-2, 2017 https://doi.org/10.5194/isprs-annals-IV-2-W4-75-2017 /W4: 75–82.

[27] S. Baselga, "TestGrids: evaluating and optimizing map projections," *Journal of Surveying Engineering*, vol. 145, no. 3, Article ID 04019004, 2019.

[28] W. Odziemczyk, "Application of simulated annealing algorithm for 3D coordinate transformation problem solution," *Open Geosciences*, vol. 12, no. 1, pp. 491–502, 2020.

[29] J. Cheng and B. Wang, "Flocking control of mobile robots with obstacle avoidance based on simulated annealing algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 7357464, 9 pages, 2020.

[30] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Springer Netherlands, Dordrecht, Netherlands, 1987.

[31] P. M. Pardalos and H. E. Romeijn, Eds., *Handbook of Global Optimization*, Springer US, Vol. Vol. 1&2, New York, NY, USA, 2002.

[32] G. Qingming and L. Jinshan, "Generalized shrunken-type robust estimation," *Journal of Surveying Engineering*, vol. 125, no. 4, pp. 177–184, 1999.

[33] A. Wieser and F. K. Brunner, "Short static GPS sessions: robust estimation results," *GPS Solutions*, vol. 5, no. 3, pp. 70–79, 2002.

[34] P. Trasak and M. Stroner, "Outlier detection efficiency in the high precision geodetic network adjustment," *Acta Geod Geoph*, vol. 49, pp. 161–175, 2014.

[35] M. Jalali-Heravi and E. Konouz, "Use of quantitative structure-property relationships in predicting the Krafft point of anionic surfactants," *Internet Electronic Journal of Molecular Design*, vol. 1, pp. 410–417, 2002.

[36] A. Janowski and J. Rapinski, "M-split estimation in laser scanning data modeling," *Journal of the Indian Society of Remote Sensing*, vol. 41, no. 1, pp. 15–19, 2013.

[37] A. Janowski, "The circle object detection with the use of Msplit estimation," *E3S Web of Conferences*, vol. 26, Article ID 00014, 2018.