Comparison of some robust parameter estimation techniques for outlier analysis applied to simulated GOCE mission data

B. Kargoll
Institute of Theoretical Geodesy (ITG),
University of Bonn, Nussallee 17, D-53115 Bonn, Germany

Abstract. Until now, methods of gravity field determination using satellite data have virtually excluded robust estimators despite the potentially disastrous effect of outliers. This paper presents computationally-feasible algorithms for Huber's M-estimator (a classic robust estimator) as well as for the class of R-estimators which have not traditionally been considered for geodetic applications. It is shown that the computational time required for the proposed algorithms is comparable to the direct method of least squares. Furthermore, a study with simulated GOCE satellite gradiometry data demonstrates that the robust gravity field solution remains almost unaffected by additive outliers. In addition, using robustly-estimated residuals proves to be more efficient at detecting outliers than using residuals resulting from least squares estimation. Finally, the non-parametric R-estimators make less assumptions about the measurement errors and produce similar results to Huber's M-estimator, making that class a viable robust alternative.

Keywords. GOCE, satellite gradiometry, robust parameter estimation, rank norm, outlier diagnostics

---

1 Introduction

The GOCE satellite mission, currently under preparation for its launch in 2006, will provide tens of millions of satellite gradiometry (SGG) observations used to recover the detailed structures of the Earth's gravity field. The global gravity field will be resolved up to degree and order 250 resulting in more than 60,000 estimated spherical harmonic coefficents. To tackle this huge adjustment problem the method of least squares has been accepted as the traditional estimator to be used. In addition to its computational feasibility, the least squares estimator produces unbiased estimates with minimal variances under certain assumptions, and comprises a unified theory including consistent variance-covariance information of the estimates, tests of model adequacy, parameter tests, and outlier tests.

However, the usual assumptions underlying the least squares procedure such as normality and outlier-freeness of the observations cannot be accepted at face value (even though the term *outlier* usually implies the association with a visibly extreme observation, it is, in the context of the current paper, understood to be any observation stemming from some *contaminating distribution* different from the main distribution of the errors, and could thus comprise the case of a blunder). Outliers, even in low numbers, are known for distorting parameter and accuracy estimates, rendering them potentially useless. The common practice of detecting outliers from least squares residuals may be fruitless due to the tendency of the estimated trend to be drawn towards extreme data values and the masking effect of multiple outliers (see for instance Rousseeuw and Leroy, 2003, p. 226 and p. 234, respectively). The traditional approach to dealing with outliers was pioneered by the work of Baarda (1968) and is based on an iterative elimination of the most prominent outlier candidates ("data snooping"). With each iteration the partially cleaned data set is readjusted, which, however, may become computationally very expensive when a huge amount of observations and a large number of potential outliers are involved.

As a remedy to this problem, the current paper investigates in robust estimators, which potentially become less affected by extreme values and are able to highlight outliers in the residuals by unmasking them. Despite such benefits, these methods apparently have been ignored in the field of gravity field determination from satellite data, mainly because of the high computational effort usually associated with robust techniques. The main purpose of this paper is to show that in fact estimates with good robustness properties may be obtained in a computational time comparable to the least squares approach. Huber's classical M-estimator using metrically Winsorized residuals (Huber, 1981, p. 179ff.) and R-estimators based on rank statistics, also mentioned by Huber (1981, p. 163) and worked out for the linear model by Hettmansperger (1984) and Hettmansperger and McKean (1998), are considered, because they are found to be particularly suitable for the huge adjustment problem encountered in the GOCE mission. As it appears that R-estimators have not been used for geodetic applications, a more detailed review of the underlying theoretical ideas is given in Sect. 2.

Two essential factors, computation time and quality of estimates, were compared between a robust estimation approach and a *direct* approach to gravity field determination from SGG data (see for instance Pail and Plank, 2002; Schuh, 1996). The outlier study investigating these factors and the corresponding results are presented in Sect. 3.

It should be noted that the intention of this paper is not to discredit the least squares approach, but to promote the use of robust estimates in a complementary way as a reference to check least squares residuals for abnormal behaviour possibly caused by undetected, masked outliers.

2 Theory and implementation

The proposed robust estimators are derived in the context of a linear model of the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ random vector of observations, $\mathbf{e}$ is an $n \times 1$ random vector of unobservable disturbances, $\mathbf{X}$ is an $n \times u$ matrix of fixed coefficients, and $\boldsymbol{\beta}$ is a $u \times 1$ vector of unknown parameters. For now let the disturbances $e_1, e_2, \ldots, e_n$ be independently, identically, but not necessarily normally distributed random variables. The parameters are estimated by minimizing some function of the residuals. In robust estimation "less rapidly increasing functions" (Huber, 1981, p. 162) are used instead of the quadratic function in least squares. Therefore, differences between M- and R-estimators are mainly determined by the choice of this function with all arising *technical* implications. However, it is seen in the following subsections that *in practice* the implementation of these two classes of estimators is very similar.

2.1 Huber's M-estimator

M-estimates are obtained by minimizing

$$Q(\mathbf{e}) = \sum_{i=1}^{n} \rho(e_i) \tag{2}$$

where ρ is a symmetric function of the residuals $e_i = y_i - x_i^T\boldsymbol{\beta}$ with x_i^T the $i-th$ row of $\mathbf{X}$. The gradient of Q with respect to $\boldsymbol{\beta}$ is given by

$$\nabla Q(\mathbf{e}) = -\mathbf{X}^T\psi(\mathbf{e}). \tag{3}$$

M-estimates are not automatically scale invariant so that $\mathbf{e}$ must be divided by some scale factor σ. Consequently, (3) becomes

$$\nabla Q(\mathbf{e}, \sigma) = -\mathbf{X}^T\psi(\mathbf{e}/\sigma). \tag{4}$$

For Huber's M-estimator $\psi(\cdot)$ is defined as

$$\psi(e_i/\sigma) := \begin{cases} e_i/\sigma & : \text{ if } |e_i| < c\sigma \\ c\,\text{sign}(e_i) & : \text{ if } |e_i| \geq c\sigma \end{cases} \tag{5}$$

where c is a constant whose value depends on the percentage of outliers in the observations (see Huber, 1981, p. 87). The values for 1% and 5% are $c \approx 2.0$ and $c \approx 1.4$, respectively. Note that the least squares estimates are obtained by setting $\psi(e_i) := e_i$ (cf. for instance Koch, 1999, Chaps. 3 and 4, as a reference of the method of least squares in linear models). With (5) the *Winsorized residuals* are defined as

$$e_i^* := \psi(e_i/\sigma)\,\sigma. \tag{6}$$

The estimates $\tilde{\boldsymbol{\beta}}$ are obtained from the Newton step

$$\tilde{\boldsymbol{\beta}}^{(k+1)} = \tilde{\boldsymbol{\beta}}^{(k)} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}^* \tag{7}$$

where the relaxation factor was set equal to 1. The scale factor may be computed from the residuals after each step by $\tilde{\sigma} = 1.483\,\text{med}_i\{|e_i^*|\}$.

2.2 Construction of R-estimators

In this section only the most basic results from the work of Jaeckel (1972), Hettmansperger (1984), and Hettmansperger and McKean (1998) are stated in order to develop a practical and intuitive approach to the material. Attention is focussed on demonstrating the similarities and differences of R-estimators to Huber's M-estimator and the method of least squares.

The goal of the commonly-used least squares estimator is to minimize the variance of the residuals $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Since few extreme values may cause an unreasonable increase in variance, Jaeckel (1972) discussed an alternative measure of variability which is less sensitive to outliers. This measure of dispersion $D(\cdot)$ is defined as

$$D(\mathbf{z}) = \sum_{i=1}^{n} a(i) z_{(i)} \tag{8}$$

where $a(1) \leq \ldots \leq a(n)$ is a nonconstant set of *scores* satisfying $\sum_{i=1}^{n} a(i) = 0$, $\mathbf{z}$ is any realvalued $n \times 1$ vector, and $z_{(i)}$ are the ordered, non-decreasing elements of $\mathbf{z}$. Now, values $1, \ldots, n$, which are denoted as the *ranks* $R(\cdot)$ of the elements of $\mathbf{z}$ are assigned to the ordered $z_{(1)}, \ldots, z_{(n)}$. Then (8) is equivalent to

$$D(\mathbf{z}) = \sum_{i=1}^{n} a(R(z_{(i)}))z_{(i)}. \quad (9)$$

Substituting the arbitrary $\mathbf{z}$ by the residuals from (1) yields a rank estimate of $\boldsymbol{\beta}$ that minimizes

$$D(\mathbf{e}) = \sum_{i=1}^{n} a(R(e_i))e_i \quad (10)$$

where $e_i = y_i - x_i^T\boldsymbol{\beta}$ with x_i^T being the $i-th$ row of $\mathbf{X}$. It is remarkable that (10) could have been defined in terms of the R pseudo norm

$$||v||_R = \sum_{i=1}^{n} a(R(e_i))\, e_i \quad (11)$$

which substitutes 'one half' of the residuals from the L_2 norm

$$||v||_{L_2} = \sum_{i=1}^{n} e_i\, e_i \quad (12)$$

by rank-transformed and re-weighted residuals

$$e_i^{**} := a(R(e_i)). \quad (13)$$

Jaeckel (1972) shows that $D(\mathbf{e})$ is a nonnegative, continuous, and convex function of $\boldsymbol{\beta}$ which attains its minimum with bounded $\boldsymbol{\beta}$ if $\mathbf{X}$ has full rank, which are also familiar properties of the Gauss-Markov model of full rank. However, in contrast to the latter, the measure of dispersion $D(\mathbf{e})$ is not a quadratic function of the residuals, but rather linear, potentially reducing the effect of outliers on the estimates (Hettmansperger, 1984, p. 233).

The partial derivatives of $D(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ exist almost everywhere with gradient

$$\nabla D(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -\mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (14)$$

where

$$\mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{X}^T\mathbf{a}(R(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})). \quad (15)$$

Setting the gradient to approximately zero yields the R normal equations

$$\mathbf{X}^T\mathbf{a}(R(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \approx 0, \quad (16)$$

which are solved by $\tilde{\boldsymbol{\beta}}$. Note that the gradient need not necessarily attain exactly zero due to its non-continuous range. The normal equations (16) cannot be solved directly, and furthermore, the dispersion function, being essentially a decreasing step function, is not ideally suited for gradient methods. Therefore, Hettmansperger and McKean (1998, p. 184) suggest constructing a Newton-type algorithm in analogy to Huber's M-estimator, based on linearization of $\mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. Let $\boldsymbol{\beta}_0$ denote the true parameters and the scale factor $\tau_\varphi = 1/\int f^2(x)dx$ where f is the density function of the disturbances. Then, according to Hettmansperger and McKean (1998, p. 162) the linearization is given by

$$\begin{aligned} \mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0) \\ &\quad -\frac{1}{\tau_\varphi}\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1) \end{aligned}$$

from which a quadratic function $Q(\cdot)$ approximating the dispersion function $D(\cdot)$ is constructed by integration as

$$\begin{aligned} Q(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= \frac{1}{2\tau_\varphi}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\mathbf{X}^T\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &\quad -(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T\mathbf{S}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0) \\ &\quad +D(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0). \end{aligned}$$

Jaeckel (1972; Lemma 1) proved that $Q(\cdot)$ is indeed a good local approximation. The estimate

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \tau_\varphi(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{a}(R(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)) \quad (17)$$

minimizes the quadratic approximation $Q(\cdot)$ and solves the linearization.

Turning attention to the practical implementation of this rank-based estimator, (17) would be computed as the first Newton step by substituting initial parameter values for the true parameters. Consequently, (17) becomes

$$\tilde{\boldsymbol{\beta}}^{(k+1)} = \tilde{\boldsymbol{\beta}}^{(k)} + \tilde{\tau}_\varphi^{(k)}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}^{**(k)}. \quad (18)$$

The scale factor $\tilde{\tau}_\varphi^{(k)}$ can be estimated from the residuals of each preceeding step. A computationally feasible estimator is derived in Hettmansperger and McKean (1998, pp. 181-184), which was used for the following simulations. The final estimate of τ_φ is used for the computation of the covariance matrix of the estimated parameters $\Sigma\{\tilde{\boldsymbol{\beta}}\} = \tilde{\tau}_\varphi^2(\mathbf{X}^T\mathbf{X})^{-1}$.

The optimal method of generating the scores $a(i) = \varphi(i/(n+1))$ through some score function $\varphi(u)$ depends on the distribution of the disturbances, which in the linear model (1) was not necessarily assumed to be Gaussian. In the current paper the score functions $\varphi(u) = \sqrt{12}(u - 1/2)$ and $\varphi(u) = \text{sign}(u - 1/2)$ generating the Wilcoxon and the sign pseudo-norm are considered for the following reasons. Hettmansperger and McKean (1998) show in theory that the Wilcoxon pseudo-norm exploits the information contained in the observations almost as efficiently as least squares when the errors are Gaussian and outlier-free, and have good robustness when outlying observations are present. The sign pseudo norm is used, because it is equivalent to the well-known L_1 norm. However, while the estimates generated by the sign pseudo-norm can be easily computed by means of a block algorithm (see Sect. 2.3), the evaluation of the L_1 norm, usually based on a simplex-type algorithm, would not be possible for the given problem as it requires that $\mathbf{X}$ be stored as one piece in the working memory.

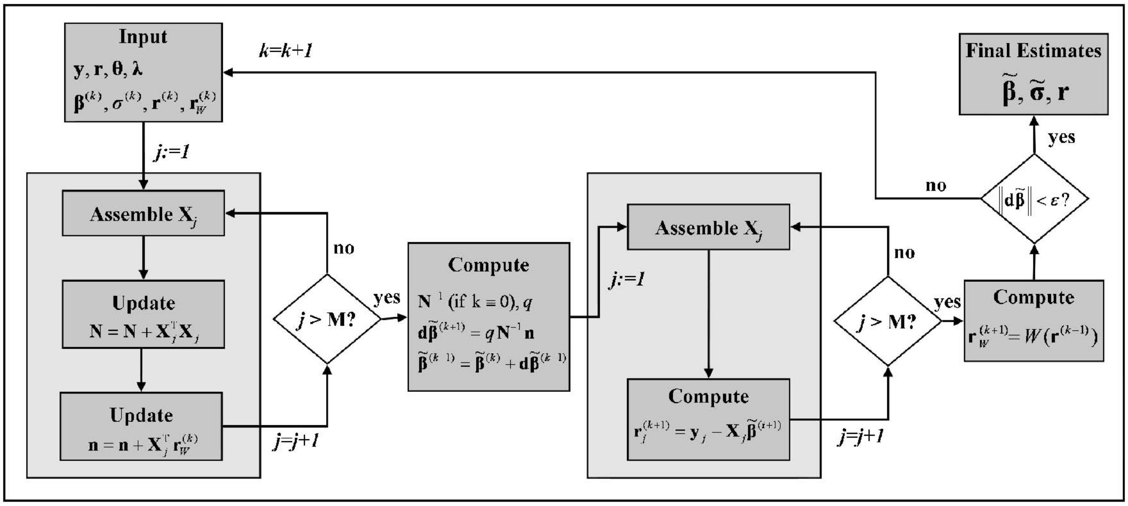

Fig. 1 Flowchart of the proposed robust gravity field solver. Input are SGG observations y_i of the three diagonal tensor components, spherical positions $(r_i, \theta_i, \lambda_i)$, and start values of the parameters $\boldsymbol{\beta}$, the scale factor σ, and the (modified) residuals generated by $\boldsymbol{\beta}_0$. $\mathbf{X}_j$ denotes the j-th block ($j = 1, \ldots, M$) of the design matrix, which is used to compute the normal equation matrix $\mathbf{N}$ and the right hand side $\mathbf{n}$ of the system $\mathbf{N}\boldsymbol{\beta} = \mathbf{n}$. $\mathbf{r}_W$ is the vector of residuals modified by a weight function $W(\cdot)$ according to (6) or (13). If the parameter update exceeds ε, the next Newton step $(k+1)$ is performed, otherwise the algorithm terminates with final estimates. The final residuals $\boldsymbol{r}$ are studentized and used for outlier detection.

2.3 Implementation of the robust gravity field solver

The functional model for the adjustment of GOCE SGG observations is obtained by taking the second derivatives of the mathematical representation of the Earth's gravitational potential

$$V(r,\theta,\lambda) = \frac{GM}{r}\left\{1 + \sum_{\ell=2}^{\ell_{max}}\sum_{m=0}^{\ell}\left(\frac{a}{r}\right)^{\ell} \times \bar{P}_{\ell m}(\cos\theta)\left(\bar{C}_{\ell m}\cos m\lambda + \bar{S}_{\ell m}\sin m\lambda\right)\right\}$$

where G denotes the geocentric gravitational constant, M the Earth's mass, and a the semi-major axis. The triple (r, θ, λ) represents the spherical coordinates of a point, ℓ and m the degree and order, ℓ_{max} the maximum degree of the expansion, and $\bar{P}_{\ell m}$ the fully normalized associated Legendre functions. The model as linear functions of the desired parameters $\bar{C}_{\ell m}$, $\bar{S}_{\ell m}$ (the fully normalized harmonic coefficients) can be expressed as the linear model (1).

As the design matrix $\mathbf{X}$ eventually contains millions of observations, it becomes far too large to be processed in one piece. Consequently, it is not possible to compute the Newton steps (see Fig. 1 for the processing flow chart) as in (7) or (17). Therefore, $\mathbf{X}$ is assembled in parts, with each part $\mathbf{X}_j$ ($j = 1, \ldots, M$) containing 750 rows. The normal equation matrix is computed within the first Newton step by $\mathbf{N} := \mathbf{X}^T\mathbf{X} = \sum_{j=1}^{M}\mathbf{X}_j{}^T\mathbf{X}_j$, and after inversion, $\mathbf{N}^{-1}$ is stored for the following steps. This procedure works for, say $\ell_{max} = 90$, as shown in the performed simulation study (see Sect. 3). To reach the GOCE mission goal of a resolution of $\ell_{max} = 250$, the algorithm must be modified, as $\mathbf{N}$ would also exceed the working memory (see *Outlook*).

The residuals generated by the parameter start values $\boldsymbol{\beta}_0$ are modified to $\mathbf{r}_W$ according to (6) or (13). Using $\mathbf{n} := \mathbf{X}^T\mathbf{r}_W = \sum_{j=1}^{M}\mathbf{X}_j{}^T\mathbf{r}_{W,j}$ and setting the relaxation parameter $q := 1$ for Huber's M-estimator, or $q := \tilde{\tau}_\varphi$, respectively, for one of the R-estimators, the parameter update $d\tilde{\boldsymbol{\beta}} = q\mathbf{N}^{-1}\mathbf{n}$ is computed. The new residuals are obtained piece by piece by assembling $\mathbf{X}$ block-wise again. In case the parameter update exceeds a prescribed ε the next Newton step is performed with updated start values. Otherwise, the current estimates are saved as the final solution. The residuals are then used for subsequent outlier analysis.

3 Simulation Study

The goal of the current simulation study is, firstly, to investigate the convergence rate of the robust estimators, because the computation time of each Newton step corresponds approximately to the entire computation time of the least squares estimation (about 4 hours on a single 3.06 GHz processor with 1 GB RAM). Secondly, the quality of the robust gravity field solutions is compared to the least squares solution. Finally, the success of outlier detection is evaluated by analyzing studentized residuals.

3.1 The Test Data

The observation functionals were computed on a sun-synchronous orbit of 23 days with an initial altitude of 250 km and an inclination of 96.6°. They were sampled equally at a rate of 4 s yielding altogether 496,430 positions and 1,489,290 values of the three main diagonal elements of the gradient tensor. The trend, computed from EGM96 coefficients up to degree and order 90, was superimposed by white noise with standard deviation $\sigma = 1$ mE. From these observations two data sets containing additional outliers (generated as realizations of uniformly distributed random variables between 3 and 50 mE) were deduced. The first set contains 1% additive outliers and the second 5%, distributed randomly over the zz-component. The observations of the xx- and the yy-component were not altered. Since the true outlier distribution will be unknown, a rather pessimistic measurement scenario was simulated by selecting the outlier ratio and bandwidth as specified above.

3.2 Results

The absolute differences between the estimated and the reference solution (least squares parameters estimated from the observations containing no outliers) were computed. Fig. 2 shows that the mean and median values over all orders of the same degree are ten times larger for the least squares estimates (LSE) than for the three robust solutions. Huber's M-estimates (HME) are equal to the Wilcoxon norm estimates (WNE), while the sign norm estimates (SNE) performs slightly worse than the WME.

Table 1 summarizes the geoid height differences between the reference solution and the estimated solutions. The differences between the reference solution and the geoid heights computed from the true EGM96 model are also given. It is seen that the robust solutions are already as close as a few millimeters to the reference values after one Newton iteration, and they converge fully after the second iteration. The least squares geoid heights differ significantly from the reference heights, and they explode for the second data set containing 5% outliers (lower part of Table 1). Using the worsened least squares estimates as start values, the robust estimates converge only after five iterations. In comparison to LSE the robustly estimated geoid heights were considerably less affected by the outliers.

Table 2 gives a summary of the performance of outlier detection by means of internally studentized residuals, defined as $rs_i = r_i/(\tilde{\sigma}\sqrt{1-h_i})$ for the least squares residuals and $rs_i = r_i/(\tilde{\sigma}\sqrt{1-Kh_i})$ for the residuals of the R-estimates. The latter are modified by K, as R-estimators do not project y orthogonally into the column space of $\mathbf{X}$ (see Hettmansperger and McKean, 1998, p. 197ff.). h_i denotes the i-th diagonal element of the orthogonal projector $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. All robust estimators detect 99.8% of the outliers (indicating a very high test power), especially 'unmasking' all outliers larger than 5 mE. The undetected, small outliers are located in the range of the measurement noise, which makes them hard to identify. By contrast, the least squares studentized residuals are much smaller (because the estimated standard deviation is inflated by the outliers), leaving even a high number of large outliers undetected.

Approximately 4% of the "good" observations were wrongly marked as outliers when using one of the robust estimators. This number could be improved only at cost of the test power, i.e. a larger number of outliers would remain undiscovered. For example, if one decreased the error number from 4% down to 0.2% by raising the threshold, one would diminish the performance rate by approximately 1%. However, for the robust estimators the choice of the actual test power is not a crucial point, because none of the observations are deleated, but their residuals downweighted.

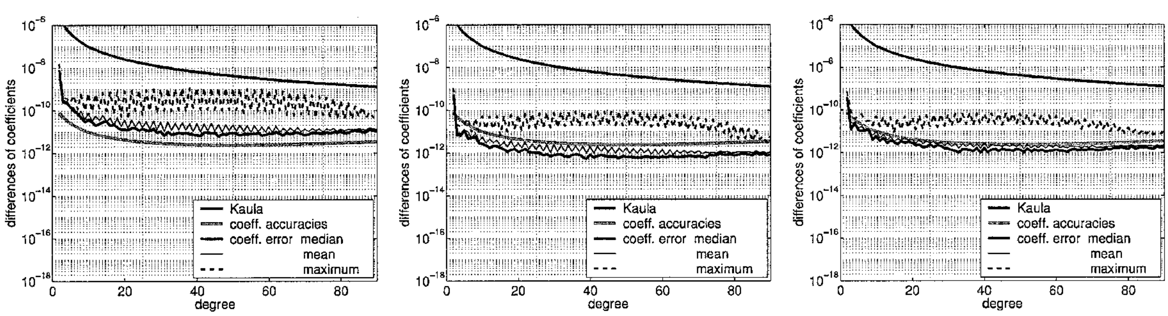

Fig. 2 Median, mean and maximum values of absolute differences between estimated (with 5% outliers) and reference (without outliers) coefficients over all orders of the same degree. From left to right: Least squares (LSE), Wilcoxon norm (WNE), and sign norm estimates (SNE) (the figure for Huber's M-estimates is the same as for the WNE and was omitted).

Table 1. Reconstruction of second-level information on a $1° \times 1°$ grid: differences between the geoid heights in meters computed from the true model (EGM96) and estimated solutions ("Reference": Least squares solution without outliers; "Least Squares": Least squares with outliers); upper part: 1% outliers, lower part: 5% outliers.

	global		$-80° \leq \phi \leq 80°$		local			
1% outliers	min	max	min	max	min	max	mean	σ
Reference	-0.012	+0.013	-0.012	+0.013	-0.008	+0.008	-0.000	0.003
Least Squares	-0.055	+0.082	-0.055	+0.077	-0.031	+0.033	-0.000	0.009
1.iteration								
Wilcoxon	-0.015	+0.015	-0.015	+0.015	-0.009	+0.011	-0.000	0.003
Sign	-0.017	+0.017	-0.017	+0.017	-0.014	+0.013	-0.000	0.004
Huber	-0.016	+0.015	-0.016	+0.015	-0.010	+0.011	-0.000	0.003
2.iteration								
Wilcoxon	-0.015	+0.015	-0.015	+0.015	-0.010	+0.010	-0.000	0.003
Sign	-0.017	+0.018	-0.017	+0.018	-0.014	+0.015	-0.000	0.004
Huber	-0.014	+0.014	-0.014	+0.014	-0.009	+0.010	-0.000	0.003
5% outliers	min	max	min	max	min	max	min	σ
Reference	-0.012	+0.013	-0.012	+0.013	-0.008	+0.008	-0.000	0.003
Least Squares	-0.236	+0.333	-0.236	+0.333	-0.111	+0.120	-0.002	0.035
1.iteration								
Wilcoxon	-0.027	+0.022	-0.027	+0.022	-0.009	+0.011	-0.000	0.003
Sign	-0.038	+0.041	-0.038	+0.041	-0.027	+0.024	-0.000	0.007
Huber	-0.060	+0.080	-0.060	+0.080	-0.032	+0.034	-0.001	0.010
5.iteration								
Wilcoxon	-0.026	+0.028	-0.026	+0.028	-0.015	+0.013	-0.000	0.004
Sign	-0.023	+0.022	-0.023	+0.021	-0.018	+0.015	-0.000	0.005
Huber	-0.026	+0.027	-0.026	+0.027	-0.015	+0.013	-0.000	0.004

Table 2. Outlier detection for the second data set containing 5% outliers. perf: percentage of correctly identified outliers (second column), error: percentage of observations wrongly marked as outliers (third column); columns 4-7: numbers of unidentified outliers of given sizes. The last row contains the distribution of the implemented outliers.

	perf	error	3-4 mE	4-5 mE	5-6 mE	>6mE
Least Squares	89.3%	0.0%	519	523	493	1,123
Huber	99.8%	4.1%	51	8	0	0
Wilcoxon	99.8%	4.3%	50	5	0	0
Sign	99.8%	4.3%	52	6	0	0
Total outliers	-	-	519	525	499	23,279

4 Discussion and Outlook

It was seen that Huber's M-estimator and the R-estimators remain robust when a small percentage of the observations are contaminated by additive outliers. Robustly estimated spherical harmonics coefficients and derived second-level products such as geoid heights became far less affected than with the least squares approach. Consequently, the "unmasked" outliers were detected almost perfectly by comparing the robustly estimated studentized residuals with a threshold value. Huber's M-estimator and the Wilcoxon norm estimator produced very similar results, and were slightly superior to the less efficient sign norm estimates (which is equivalent to the L_1 norm). All robust estimates converged after a few iterations when heavily distorted least squares start values were used. When valid *a priori* information was used, they converged within one step, i.e. the computational effort was essentially the same as for computing the least squares solution, making robust procedures feasible.

For the future it is intended to robustly estimate models up to degree and order 250 (the planned resolution of the GOCE mission). Since the proposed algorithm allows the block-wise processing of the normal equations, this can be easily accomplished by implementation on a parallel computer system.

Acknowledgments

The support by BMBF through the GOCE-GRAND project within the "Geotechnologien-Programm" is gratefully acknowledged.

References

Baarda, W. (1968). *A testing procedure for use in geodetic networks.* Publications on Geodesy by the Netherlands Geodetic Commission 2(5).

Hettmansperger, T.P. (1984). *Statistical inference based on ranks.* John Wiley, New York.

Hettmansperger, T.P. and J.W. McKean (1998). *Robust nonparametric statistical methods.* Arnold, London.

Huber, P.J. (1981). *Robust Statistics.* John Wiley, New York.

Jaeckel, L.A. (1972). *Estimating regression coefficients by minimizing the dispersion of the residuals.* The Annals of Mathematical Statistics 43(5):1449-1458.

Koch, K.-R. (1999). *Parameter estimation and hypothesis testing in linear models.* Springer, Berlin/Heidelberg.

Pail, R. and G. Plank (2002). *Assessment of three numerical solution strategies for gravity field recovery from GOCE satellite gradiometry implemented on a parallel platform.* Journal of Geodesy 76:462-474.

Rousseeuw, P.J. and A.M. Leroy (2003). *Robust regression and outlier detection.* John Wiley, New York.

Schuh, W.D. (1996). *Tailored numerical solution strategies for the global determination of the Earth's gravity field.* Mitteilungen der geodaetischen Institute der TU Graz 81.