Comparison of outlier detection algorithms for GOCE gravity gradients

J. Bouman (1), M. Kern (2), R. Koop (1), R. Pail (2), R. Haagmans (3), T. Preimesberger (4)
(1) SRON National Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands
(2) Institute of Navigation and Satellite Geodesy, TU Graz, Steyrergasse 30, 8010 Graz, Austria
(3) Science and Applications Department, ESA/ESTEC, Keplerlaan 1, 2200 AG Noordwijk, The Netherlands
(4) Austrian Academy of Sciences, Space Research Institute, Schmiedlstrasse 6, 8042 Graz, Austria

Abstract. GOCE will be the first satellite ever to measure the second derivatives of the Earth's gravitational potential in space. It will be possible to derive a high accuracy and high resolution model of the gravitational field if systematic errors and/or outliers have been removed from the data. It is necessary to detect outliers in the data pre-processing because undetected outliers may lead to erroneous results when the data are further processed, for example in the recovery of a gravity field model. Outliers in the GOCE gravity gradients will be searched for and detected in the gravity field analysis pre-processing step.

In this paper, a number of algorithms are discussed that detect outliers in the diagonal gravity gradients. One of them combines wavelets with either a statistical method or filtered gradients with an identification rate of about 90% or more. Another high performing algorithm is the combination of three methods, that is, the tracelessness condition (a physical property of the diagonal gradients), comparison with model or filtered gradients, and along-track interpolation of gradient anomalies. Using two sets of simulated gravity gradients, the algorithms are compared in terms of their identification rate and number of falsly detected outliers. In addition, it is shown that the quality of the gravity field solution is very much affected by outliers. Undetected outliers can degrade the gravity field solution by up to twenty times as compared with a solution without outliers.

Keywords. Outliers · GOCE mission · Gradiometry · Statistical tests

1 Introduction

The main goal of the GOCE mission (expected launch in August 2006) is to provide unique models of the Earth's gravity field and of its equipotential surface, as represented by the geoid, on a global scale with an accuracy of 1 cm at 100 km resolution (ESA 1999). To this end, GOCE will be equipped with a GPS receiver for high-low satellite-to-satellite tracking (SST-hl), and with a gradiometer for observation of the gravity gradients (GG). Only the latter will be considered in this paper.

Even after the in-flight calibration, the observations will be contaminated with stochastic and systematic errors. Systematic errors include GG scale factor errors and biases (Cesare 2002) which are corrected for in the external calibration step (see e.g. Arabelos and Tscherning 1998; Bouman et al. 2004). In addition, outliers in the GOCE gravity gradients are searched for and detected in the gravity field analysis (GFA) pre-processing step. If some remain undetected, they may seriously affect the accuracy of the final GOCE gravity field model (Kern et al. 2004).

A vast number of outlier detection methods exists and a selection is discussed in this paper. Outliers are searched for in simulated gravity gradient time series contaminated with noise and outliers. The performance of the methods is evaluated with respect to the detection rate and the type I error (rejecting correct data). Section 2 details several outlier detection methods and section 3 shows numerical examples. One alternative for the time-wise methods studied here, is presented by (Tscherning 1991).

2 Outlier detection methods

2.1 TSOFT outlier detection algorithm

The TSOFT outlier detection algorithm is based on the algorithm presented by Vauterin and Van Camp (2004). The idea is to low-pass filter the gravity gradient time series which tends to reduce the outliers. If for certain points the difference between the filtered and unfiltered time series is above a certain threshold, thr_1, then these points are likely to be outliers. The effect of the low-pass filter is not only a reduction of the size of the outliers, but also a redistribution of the power over neighbouring points. In addition, an outlier that is close to another outlier may mask that

outlier. Therefore, iteration is necessary, replacing the detected outliers in the original time series by the filtered values. This updated time series is then again low-pass filtered and again outliers can be searched for, etc. The final low-pass filtered time series, with most outliers removed, is tested against the original time series with outliers. If the difference is above a threshold, $thr_2 \leq thr_1$, then an outlier is detected. The second threshold can be smaller than the first one since the final low-pass filtered time series is affected less by outliers than the filtered times series in the iteration. The thresholds themselves have to be determined using simulated data or by trial and error.

2.2 Wavelet outlier detection algorithm

Single and higher level Haar wavelet can be used to detect outliers. The wavelet outlier detection is explicitely explained in the paper by Kern et al. (2004). It searches for discontinuities in the signal. A single level outlier detection algorithm may be formulated as follows.

[1] Compute detailed and smoothed wavelet coefficients using the forward wavelet transformation ($k = 1, \ldots, m/2 - 1$)

$$\begin{aligned} s_{1,k} &= (x_{2k} + x_{2k+1})/\sqrt{2}, \\ d_{1,k} &= (x_{2k} - x_{2k+1})/\sqrt{2}, \end{aligned} \tag{1}$$

with data vector $x_i = \{x_1 \ldots x_m\}$.

[2] Threshold the detailed coefficients by setting a threshold td

$$d'_{1,k} = \begin{cases} d_{1,k} & \text{for } |d_{1,k}| < td, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

[3] Reconstruct the signal

$$\begin{aligned} x^w_{2k} &= (s_{1,k} + d'_{1,k})/\sqrt{2}, \\ x^w_{2k+1} &= (s_{1,k} - d'_{1,k})/\sqrt{2}. \end{aligned} \tag{3}$$

[4] Compute the residual signal using the reconstructed signal x^w

$$r_i = x_i - x^w_i, i = 1, \ldots, m. \tag{4}$$

[5] Apply a pattern recognition program on the residuals to identify the position of the outliers.

2.3 Dixon test

The Dixon test is a hypothesis test that uses the ratio of differences between a possible outlier and its nearest or next-nearest neighbour (data excess) to the range. The data have to be normally distributed. Let the data vector $x_i = \{x_1 \ldots x_m\}$ be sorted in ascending order and let $t_{1-\alpha}(m-1)$ be the Student's t-distribution, which depends on the significance level $\alpha \in (0, 1)$ and the number of observations m. An α-outlier region for upper outliers is defined as

$$\text{out}(\alpha, m) := \{r_j > t_{1-\alpha}(m-1); j = 1, 2\} \tag{5}$$

where the test functions are given as (Barnett and Lewis 1994)

$$r_1 = \frac{x_m - x_{m-1}}{x_m - x_2}, \; r_2 = \frac{x_m - x_{m-1}}{x_m - x_3}. \tag{6}$$

If one of the test functions r_j exceeds the critical value $t_{1-\alpha}(m-1)$, the largest observation is an outlier or the distribution is not normal. The test statistic r_1 does not contain the smallest value x_1 to avoid masking effects (large denominator). Similarly, the test statistic r_2 can be used to avoid masking effects from the smallest two values (x_1 and x_2). Because the Dixon test is a very robust method, one may expect that it also works in the presence of data gaps. This was, however, not investigated in this paper.

2.4 Tracelessness condition

The sum of the diagonal gravity gradients, also called Laplace's equation or tracelessness condition, has to be zero, which is a physical property of the gravity gradients. The gravitational potential is a harmonic function outside the attracting masses (Heiskanen and Moritz 1967). However, before external calibration, the gradients suffer from systematic errors of which a bias and scale factor errors are the most important. The effect of a scale factor error is the largest at a frequency of 0 Hz or the mean value. Of course, also the bias is manifest at this frequency. Therefore, the following condition equation is considered

$$E\{V_{xx} + V_{yy} + V_{zz}\} - \text{ median } = 0 \tag{7}$$

where E is the expectation operator and the median is the median of the point-wise Laplace's equation of the time series considered. Note that in the GOCE case the rotational terms, caused by rotation of the satellite, have been removed as good as possible from the gradients using the differential accelerations (Cesare 2002). The w-test is used, i.e., if the tracelessness condition is violated then an outlier is detected. The trace is weighted with the a priori error of the GOCE gravity gradients neglecting along-track error correlations (Bouman 2004). The major drawback of the tracelessness condition is that the outlier detection is ambiguous, i.e., one cannot discriminate between outliers on V_{xx}, V_{yy} and V_{zz}. The advantage

is that it is a sensitive method. The smaller the signal-to-noise-ratio (SNR), the easier it is to detect outliers. In fact, the SNR can not be smaller since the sum of the diagonal gradients should be zero.

2.5 Gravity gradient anomalies

The GOCE gravity gradients could be confronted with gravity gradients generated from a global Earth gravity field model. If the difference between the two, weighted with the sum of the respective errors, is above a certain threshold, then an outlier is detected (the median of differences is subtracted to account for the GOCE gravity gradient bias and scale factor error). More details on this w-test are given in Bouman (2004). The advantage of this method with gravity gradient anomalies is that all gradients can be tested separately and point-wise. A disadvantage is that the accuracy of the model gradients may be low compared to the GOCE gradients, which makes this a less sensitive method. In addition, the two sets of gradients have different measurement bandwiths, which may restrict the test.

Alternatively, one could consider the along-track interpolation of GG anomalies. An anomaly at time $t = t_i$ is compared with the predicted anomaly at time $t = t_i$, where the prediction is based on anomalies at $t = t_j, j \neq i$. Many interpolation methods could be used; splines are used here since they are simple and fast and the interpolation errors are small (Bouman 2004). The advantage of the along-track interpolation is that the gradients can be tested separately, but several points are combined which may lead to masking effects, that is, outliers close to each other can not be well separated.

2.6 Combination solutions

One possibility to improve the results is to combine two or more of the methods described above. The combination of the TSOFT algorithm and the wavelet method are considered, while the latter is also combined with the Dixon test. A data point is flagged as an outlier if it is detected in both methods.

Also considered is the combination of the tracelessness condition, gravity gradient differences and the interpolation of these differences. Since the tracelessness condition is a sensitive but ambiguous method, the other two methods are used to confirm a detected outlier by the tracelessness condition. In other words, if an outlier on V_{xx}, V_{yy} and V_{zz} is detected by the tracelessness condition and this outlier is confirmed by either the gradient differences or the interpolation, a data point is flagged as an outlier.

2.7 Other methods

Besides the above methods, the so-called m-estimator (Mayer 2003) and two thresholding methods (Kern et al. 2004) were studied. The m-estimator has a high outlier detection rate, but it also has a large type I error, that is, up to one out of five observations is erroneously detected as an outlier in the tests made. This method will therefore not be discussed. The threshold methods detect an outlier if the difference between the value at a given data point and the mean or median is above a certain threshold. The threshold may be linked to the standard deviation of the data or to some fixed value. These methods, however, suffer from a relatively large type I error while the number of detected outliers is relatively small. Therefore, these methods are not considered here.

3 Numerical results

Two data sets with different characteristics were studied. One is a small data set with a length of 1 day which contains various types of outliers. The second data set has a length of 59 days and contains single and bulk outliers. This set allows for gravity field analysis.

3.1 Small data set with various outliers

The first data set used in this study consists of the diagonal gravity gradients V_{xx}, V_{yy} and V_{zz} which were simulated using EGM96 (Lemoine et al. 1998) for a 1 day orbit with a sampling rate of 1 s. Simulated, correlated noise was added to the signals, the data statistics are given in Table 1. The model gradients which are required for some methods were generated using OSU91A (Rapp et al. 1991). A first test was done that used the noisy gradients without any outliers (case 1a). The type I error is (close to) zero as one would hope. However, this is not to be expected for the tracelessness condition, model gradients and spline interpolation. These all use the w-test with a critical value of $k = 2$, which would mean that approximately 4.6% of the observations is rejected although they are correct. For the tracelessness condition and spline interpolation, however, the type I error is 0%. This may be due to the error correlation between the simulated gradients which is neglected. The model gradients have a larger type I error but this is dominated by the model error, that is, the difference between EGM96 and OSU91A. The type I error is probably larger than expected because we have used a simple scale unit matrix as error covariance matrix.

Table 1 Noise, outlier and anomaly properties, values in [mE].

Small set	(86,351 pts)	V_{xx}	V_{yy}	V_{zz}
noise	mean	1443.7	-805.2	2248.9
	rms	2.2	4.4	5.7
outliers	mean	0.5	0.3	0.0
	rms	58.9	27.9	52.7
	number	3,891	420	1,988
anomalies	mean	0.0	1.5	-1.5
	rms	36.4	35.3	58.9
Large set	(5,097,835 pts)	V_{xx}	V_{yy}	V_{zz}
noise	mean	0.0	0.0	0.0
	rms	10.1	2.7	10.0
outliers	mean	0.0	0.0	0.0
	rms	78.5	78.5	78.5
	number	83,153	83,153	83,153
anomalies	mean	0.0	-0.4	0.4
	rms	37.2	35.3	60.0

Table 2 Type I error for case 1a (no outliers, small data set); TS – TSOFT algorithm, W – wavelet detection, TSW – TSOFT + wavelet, TR – tracelessness condition, M – model gradients, SP – spline interpolation, TMS – TR + M + SP.

Method	V_{xx}	V_{yy}	V_{zz}
TS	0.1%	0.1%	0.1%
W	0.0%	0%	0.2%
TSW	0.2%	0.1%	0.4%
TR	0.0%	0.0%	0.0%
M	6.2%	6.0%	5.9%
SP	0%	0%	0%
TMS	0%	0%	0%

A second test was done with outliers on all three gradients with an absolute size varying between 0.07 E and 0.1 E (case 1b). The outliers on V_{xx} are randomly distributed single outliers. The outliers on the V_{yy} component are an offset of 0.5 E during one minute ($t = 20 - 79$ s) and a bulk of outliers during six minutes ($t = 5000 - 50359$ s). Finally, the outliers on the V_{zz} component consist of randomly distributed 'twangs', i.e., outliers at $t = t$ that are followed by an other outlier of opposite sign and of the same size at $t = t + 1$. In total there are 3891, 420 and 1988 outliers on the V_{xx}, V_{yy} and V_{zz} component respectively, see also Table 1.

Outlier detection results are shown in Table 3. Rows 1-3 show the TSOFT algorithm (TS), wavelets (W) and their combination (TSW) respectively. Rows 4 and 5 show the tracelessness condition (TR) and the model gradients (M). Row 6, TS1, shows the results for filtered gradients, that is, the GG with outliers were filtered and these were used as model gradients to compute GG anomalies. A 2nd order low-pass Butterworth filter with a cut-off frequency of 0.2 Hz was used. The acronym TS1 is used because this method is similar to one step of the TSOFT algorithm, although the low-pass filter is different. The last three rows show respectively the results for spline interpolation (SP), the combination of the tracelessness condition, model gradients, splines (TMS) and the combination of the tracelessness condition, filtered gradients and splines (TFS).

Most of the single outliers V_{xx} are detected by all methods (we consider percentages above 95% to be good). The type I error is large for filtered gradients and splines because outliers are spread out over several data points in both methods (we consider type I errors above 5% to be too large). The 'twangs' on V_{zz} are also detected by most methods. The wavelet detection algorithm has problems because it takes the difference between two consecutive data points. The detection rate for model gradients is somewhat low due to the larger GOCE GG error and the larger difference between the 'true' GG and the model GG (OSU91A and EGM96). The offset on V_{yy} causes problems for many methods. The filter methods, TS and TS1, as well as spline interpolation fail to detect the offset. An offset tends to cancel in these methods. The wavelet outlier detection rate is low as it not only fails to identify the offset, but it also has problems with the bulk outliers. In general, the combination of different outlier detection methods gives a higher detection rate and a low type I error. The detection of the offset remains a problem, also in the combination solutions, with the exception of the combination with model gradients which detects most of the V_{yy} outliers.

3.2 Large data set and gravity field retrieval

The second data set used in this study also consists of the diagonal gravity gradients V_{xx}, V_{yy} and V_{zz} which were simulated using OSU91A for a 59 day orbit with a sampling rate of 1 s (over half a million data points). Simulated, correlated noise was added to the signals. (A test with no outliers gives roughly the same percentage of type I errors as for the small data set except for the tracelessness condition which has a type I error of 4.7%. The simulated GG errors for the large data set show no correlation between the different GG.) In addition, outliers were added to all three gradients with an absolute size varying between 0.05 E and 1.8051 E (case 2). The outliers were randomly distributed single outliers as well as bulk outliers, see Table 1 for data statistics.

Besides the detection methods discussed before, the combination of wavelets and the Dixon test was

Table 3 Detected outliers for case 1b (outliers on all three diagonal gradients, small data set); TS – TSOFT algorithm, W – wavelet detection, TSW – TSOFT + wavelet, TR – tracelessness condition, M – model gradients, TS1 – filtered gradients, SP – spline interpolation, TMS – TR + M + SP, TFS – TR + TS1 + SP.

	V_{xx}		V_{yy}		V_{zz}	
Method	correct	type I	correct	type I	correct	type I
TS	99.9%	1.3%	83.6%	0.1%	100%	0.1%
W	95.8%	0.1%	46.9%	0.0%	62.8%	0.7%
TSW	100%	3.0%	87.6%	0.1%	100%	0.9%
TR	99.9%	2.6%	99.8%	6.7%	99.9%	4.8%
M	93.6%	5.9%	92.6%	6.0%	76.9%	5.8%
TS1	99.8%	23.5%	84.5%	0.0%	100%	2.2%
SP	98.9%	11.6%	77.6%	0.0%	99.5%	2.0%
TMS	99.8%	0.5%	98.6%	0.4%	99.7%	0.4%
TFS	99.9%	0.7%	87.1%	0%	99.9%	0.1%

Table 4 Detected outliers for case 2 (outliers on all three diagonal gradients, large data set); TS – TSOFT algorithm, W – wavelet detection, TSW – TSOFT + wavelet, WD – wavelet + Dixon test, WDQL – wavelet + Dixon + QL-GFA, TR – tracelessness condition, M – model gradients, TS1 – filtered gradients, SP – spline interpolation, TMS – TR + M + SP, TFS – TR + TS1 + SP.

	V_{xx}		V_{yy}		V_{zz}	
Method	correct	type I	correct	type I	correct	type I
TS	99.0%	4.4%	99.0%	4.3%	99.0%	4.4%
W	86.4%	0.1%	86.7%	0.1%	85.6%	0.3%
TSW	99.6%	6.5%	99.6%	6.3%	99.6%	6.7%
WD	97.0%	0.0%	98.4%	0.0%	96.2%	0.1%
WDQL	100.0%	0.1%	100.0%	0.0%	100.0%	0.1%
TR	99.8%	7.7%	99.9%	7.7%	99.9%	7.7%
M	96.8%	5.9%	97.5%	6.2%	92.3%	5.9%
TS1	99.0%	5.6%	99.7%	10.5%	98.9%	5.6%
SP	86.8%	2.2%	98.6%	4.7%	86.8%	2.2%
TMS	97.7%	0.6%	99.8%	0.8%	94.6%	0.6%
TFS	99.6%	0.4%	99.9%	0.8%	99.6%	0.4%

added (WD). The cleaned GG from this method are used in Quick-Look Gravity Field Analysis (QL-GFA) to compute a global gravity field model (Pail and Preimesberger 2003). This gravity field model is used to compute GG along the orbit. Then, an additional search in the residuals between these GG and the observed GG is done in an iterative manner (WDQL).

As with the small data set, wavelets perform worse for bulk outliers (row W of Table 4). The detection rates for V_{xx} and V_{zz} are lower than for V_{yy} using splines because of the higher noise level of the former two. The combination algorithms detect almost all outliers while the type I error is small. One exception is the combination of TSOFT and wavelets, which has a large type I error. The best results are obtained by the wavelet-Dixon method in QL-GFA. Almost all outliers are detected, while the type I error is very small. The major advantage of the GFA is that the data are not only 'compared' along track or point-wise, which is the drawback of the other pre-processing algorithms considered here, but that the least-squares adjustment combines all observations.

The effect of undetected outliers can be disastrous, see Table 5. Although only 1.6% of the observations contain outliers, the gravity field solution has a very low accuracy if the outliers are not removed. Shown are the gravity anomaly differences between OSU91A and a QL-GFA solution up to degree and order 250. The error standard deviation is twenty times as high compared to a solution where no outliers are present (126.0 mGal and 6.7 mGal respectively). The wavelet - Dixon combination gives a considerable improvement compared to no outlier detection, see Table 5 and Fig. 1. It does not, however, detect all bulk outliers, which cause a visible track (Fig. 1). The best combination solution that uses pre-processing only (TFS) gives a small gravity anomaly difference (9.8 mGal). Finally, the wavelet - Dixon combination in the GFA (WDQL) gives a gravity field anomaly error which is almost at the level of no outliers (7.0 mGal), see again Table 5.

Table 5 Gravity anomaly error for case 2 (large data set); difference between OSU91A and QL-GFA up to degree and order 250, excluding polar caps of 10°.

Method	error rms [mGal]
no outliers	6.7
all outliers	126.0
WD	24.9
TFS	9.8
WDQL	7.0

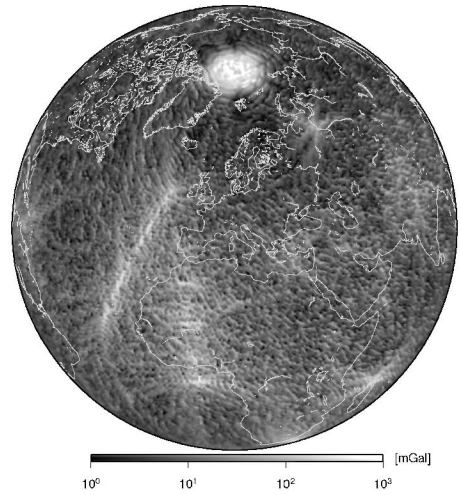

Fig. 1 Gravity anomaly differences OSU91A – QL-GFA (log scale), pre-processing outlier detection using WD.

4 Conclusions and outlook

Several outlier detection algorithms have been compared. While single outliers and 'twangs' can be detected at high rates, bulk outliers and offsets cause problems in almost all methods. Generally, a combination of methods improves the detection results. In particular, the combinations WD and TFS yield highest detection rates while having a small type I error. After applying the pre-processing methods, the overall rms of the gravity field solution can be reduced by an additional search inside the gravity field solver.

The results for the model gradients may improve as more accurate gravity models become available. Especially at the time GOCE flies, preliminary GOCE gravity field models could be used. Future studies may include orbit errors, various GG error scenarios, uncertainties in the GG a priori error model, etc.

5 Acknowledgements

Financial support for the second author came from an external ESA fellowship. Financial support for the fourth author came from the ASA contract ASAP-CO-008/03. This study was performed in the framework of the ESA project GOCE High-level Processing Facility (No. 18308/04/NL/MM). All this is gratefully acknowledged. Also the remarks by two anonymous referees are acknowledged.

References

Arabelos D, Tscherning CC (1998) Calibration of satellite gradiometer data aided by ground gravity data. Journal of Geodesy. 72: 617-625

Barnett V, Lewis T (1994) Outliers in statistical data. 3rd edition, John Wiley and Sons, Chichester, New York

Bouman J (2004) Quick-look outlier detection for GOCE gravity gradients. Paper presented at the IAG Porto meeting (GGSM 2004)

Bouman J, Koop R, Tscherning CC, Visser P (2004) Calibration of GOCE SGG data using high-low SST, terrestrial gravity data and global gravity field models. Journal of Geodesy, 78, DOI 10.1007/s00190-004-0382-5

Cesare S (2002) Performance requirements and budgets for the gradiometric mission. Issue 2 GO-TN-AI-0027, Preliminary Design Review, Alenia, Turin

ESA (1999) Gravity field and steady-state ocean circulation mission. Reports for mission selection. The four candidate Earth explorer core missions. ESA SP-1233(1). European Space Agency, Noordwijk

Heiskanen W, Moritz H (1967) Physical Geodesy. W.H. Freeman and Company, San Francisco

Kern M, Preimesberger T, Allesch M, Pail R, Bouman J, Koop R (2004) Outlier detection algorithms and their performance in GOCE gravity field processing. Accepted for publication in Journal of Geodesy

Lemoine F, Kenyon S, Factor J, Trimmer R, Pavlis N, Chinn D, Cox C, Klosko S, Luthcke S, Torrence M, Wang Y, Williamson R, Pavlis E, Rapp R, Olson T (1998) The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96. TP 1998-206861, NASA Goddard Space Flight Center, Greenbelt

Mayer C (2003) Wavelet modelling of ionospheric currents and induced magnetic fields from satellite data. PhD thesis. Geomathematics Group, Department of Mathematics, University of Kaiserslautern, Germany

Pail R, Preimesberger T (2003) GOCE quick-look gravity solution: application of the semianalytic approach in the case of data gaps and non-repeat orbits. Studia geophysica et geodaetica. 47:435-453

Rapp R, Wang Y, Pavlis N (1991) The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models. Rep 410, Department of Geodetic Science and Surveying, The Ohio State University, Columbus

Tscherning CC (1991) The use of optimal estimation for gross-error detection in databases of spatially correlated data. Bulletin d'Information, no. 68, 79-89, BGI

Vauterin P, van Camp M (2004) TSoft Manual. Version 2.0.14, Royal Observatory of Belgium, Bruxelles, Belgium. Available at http://www.astro.oma.be/SEISMO/TSOFT/tsoft.html