Proc. Kon. Ned. Akad. v. Wetensch. 98 (2), 000–000 June 26, 1995

Determining the shape of the Earth*

by P.J.G. Teunissen, M. van Gelderen, E.J. de Min and R.H.N. Haagmans

Faculty of Geodetic Engineering, Delft University of Technology, Thijsseweg 11, 2629 JA Delft, the Netherlands

Communicated by Prof. J.E. van Hinte at the meeting of September 26, 1994

1. INTRODUCTION

Geodesy is the discipline that deals with the measurement and representation of the earth surface and its external gravity field in a three-dimensional time-varying space. More specific, the two major goals of geodesy are:

1. Geometric Geodesy: The determination of the size and shape of the earth's surface S through the establishment and maintenance of regional, continental and global 1-, 2- and 3-dimensional geodetic control and densification networks.

2. Gravimetric Geodesy: The determination of the earth's gravity potential P through the solution of the various geodetic boundary value problems, thus enabling one to infer the size and shape of the geoid.

Since observed geodetic functionals are in general dependent on both geometric parameters as well as on geopotential parameters, a simultaneous determination of both S and P becomes possible in principle. These functionals are observed with terrestrial and/or space-based measurement systems. Typical examples of observed functionals are: potential differences through levelling, scalar gravity, astronomical latitude and longitude, terrestrial and/or space-

* This paper was presented in a Symposium, organized by the Council of Earth Sciences of the Royal Dutch Academy of Sciences, with the title: 'Variation in and on the Earth'. The Symposium, convened by prof. Dr. Ir. J. Bouma of the Agricultural University, Wageningen, The Netherlands, took place at the Academy in Amsterdam on November 5, 1993.

based gravity gradients and three-dimensional geocentric positions derived from angular and/or distance measurements.

The geodetic mappings from the parameter space into the observational space can be characterized as being overdetermined, underdetermined and nonlinear at the same time. Nonlinearity is a consequence of the existing non-homogeneity between the various observational functionals. Overdeterminancy is preferably introduced so as to be able to statistically test hypothesized alternative models. And underdeterminancy is present as a consequence of the observational sample being finite.

2. GEOPOTENTIAL MODELLING

Underdeterminancy and instability are particularly present in case of geopotential modelling. The procedure for determining an approximation to the continuous geopotential field P is briefly as follows. Based on the working hypothesis that mass in the earth's exterior is absent (or reduced for) and the geodetic functionals are observed at each point of S, the observation equations of gravimetric geodesy can be modelled as nonlinear, free and oblique boundary value problems. The term 'free' implies that the observed functionals are given on the boundary S which is also assumed to be unknown. This is a significant difference with ordinary boundary value problems (e.g. Dirichlet, Neumann, Robin), for which the functionals apply to a given boundary. In addition to being free, the geodetic boundary value problem is also of the 'oblique' type, because it is formulated along the direction of gravity, which is oblique with respect to the normal to the surface S. An approximate solution of P is obtained after three levels of simplifications (linearization, spherical approximation, and constant radius approximation). The solution is then analytically expressible in terms of integrals over the sphere of the convolution type, such as the Stokes, Vening–Meinesz, or Hotine integrals. These analytical solutions are however based on the assumption that a continuous covering of S with observed functionals is available. Since this is not the case in practice, underdeterminancy results as a consequence and precautions have to be taken to allow one to come up with an estimate of the continuous geopotential field.

Least-squares collocation is a mathematical technique for estimating the earth's figure and gravitational field in case of underdeterminancy. Least-squares collocation has started from the subject of interpolation of gravity anomalies using the least-squares principle and has been generalized to heterogeneous data of different kinds. The same formulas may be interpreted in very different ways: as a statistical estimation method combining least-squares adjustment and least-squares prediction, as the solution of a geophysical inverse problem, and as an analytical approximation to the earth's potential by means of harmonic functions.

In the following three different examples of geopotential modelling based on least-squares collocation are given. They are: local gravimetric geoid determination, regional geoid determination based on satellite altimetry and global geopotential modelling based on satellite tracking data. The abstract is concluded with a brief discussion on the applicability of the method of least-squares collocation.

3. LOCAL GRAVIMETRIC GEOID DETERMINATION

One example of applying least-squares collocation in physical geodesy is the computation of geoid heights from gravity anomalies. The basic formula is Stokes' integral formula, which relates the geoid height in one point P to all gravity anomalies covering the entire surface of the earth. In practice only discrete measurements of gravity are given. Based on these discrete measurements a continuous function must be predicted which is used to compute geoid heights by Stokes' integral formula. Least-squares collocation combines these two steps. Implicitly least-squares prediction is applied to the given discrete data to predict gravity values covering the entire surface of the earth, followed by Stokes' integration to obtain the geoid height contribution in the desired points.

The central part in the method is the signal covariance function which describes the covariances between gravity signal values in different points. The Stokes' integral is a linear operator, which is applied to the signal auto-covariance function of gravity anomalies to obtain the cross covariance function between geoid heights and gravity anomalies. This linear operator is usually rather difficult in the space domain, but simple in the spherical harmonic spectral domain. In this way, an easy evaluation formula is obtained: input are the discrete measurements, output are the desired continuous results. No explicit predictions or evaluations of integral formulas are needed.

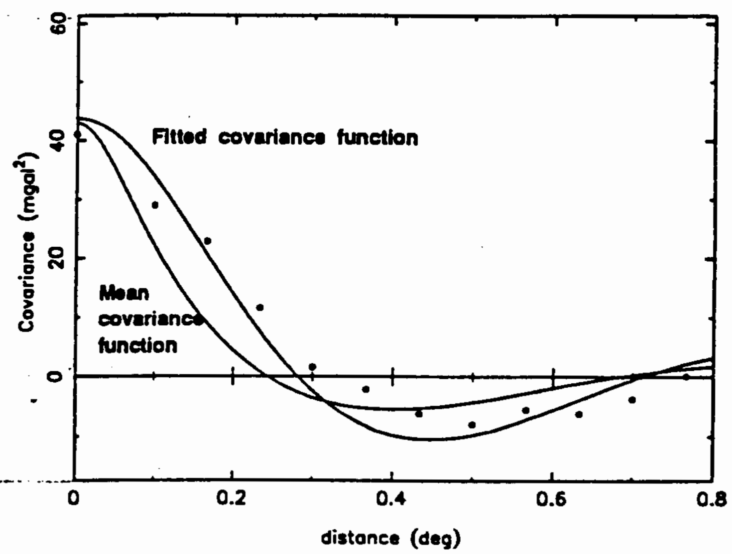

Fig. 1. Empirical covariance function values (dots) and two analytical ones.

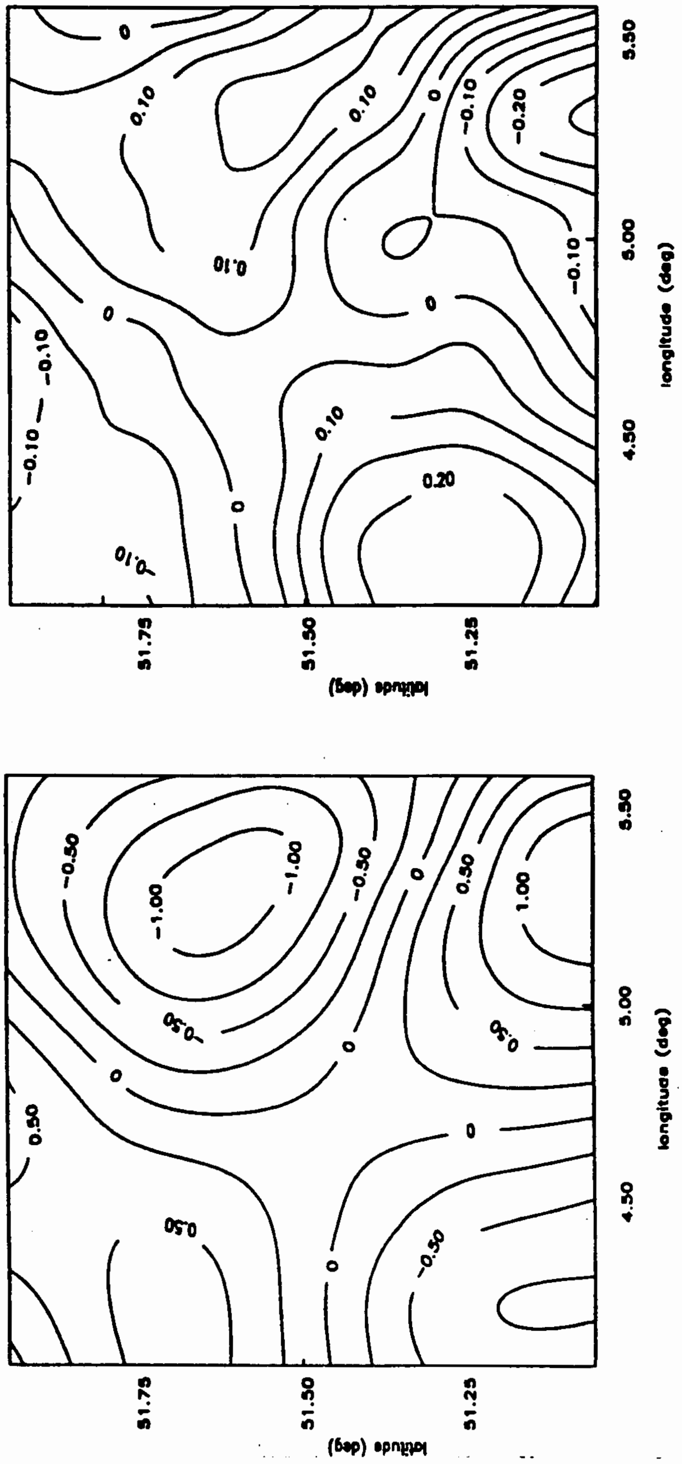

Fig. 2. Geoid results (in m) from fitted covariance function and differences from using another covariance function.

One remaining question is which covariance functions to use? One may expect that another covariance function will give other results. From the given discrete gravity data an empirical covariance function can be computed. Then, one can try to fit an analytical model to describe this empirical covariance function as good as possible. In geodetic literature it is not unusual to use a mean model for the entire earth (Basic, 1989).

Now a practical example is given. For the area in and around the Netherlands mean gravity anomalies are given. From these anomalies an empirical covariance function is computed and an analytical model is fitted. Together with the mean earth model they are given in fig. 1. From these gravity data and two analytical covariance function models the geoid contribution in part of the Netherlands is computed twice. Figure 2 shows the geoid results from the fitted model and the difference between the results using the two different covariance functions. It can easily be seen that the choice of the covariance functions is of important influence on the final geoid results.

In conclusion, it can be said that collocation is a method that is easily implemented, although on the other hand a matrix of the size of the number of data points has to be inverted, which makes it numerically unattractive. If the data density is high this matrix can even become singular. The effect in the geoid results by using another covariance function is mainly caused by the extrapolation of gravity values outside the original data area, which is automatically done by collocation.

4. SATELLITE ALTIMETRY FOR GEOID AND GRAVITY FIELD DETERMINATION

Satellite radar altimetry provides us with instantaneous sea level heights of uniform quality (around 7–10 cm) covering most parts of the worlds oceans. Analysis of altimeter data yields oceanographic, geodetic and geophysical information, such as the flow of main currents like the Gulf Stream, eddie fluxes, geoid heights and gravity anomalies. In this paragraph we will shortly focus on the geodetic and geophysical results from altimetry.

Under idealized circumstances, i.e. without external forces, the sea surface coincides with the geoid, meaning that no water flows from one point to another. Geoid heights range from −100 m to 85 m. These heights refer to a rotating reference ellipsoid with a homogeneous mass distribution which is a first approximation of the earth. In reality local and regional inhomogeneous density variations in the earth's interior force the actual equipotential or geoid to differ from the ellipsoidal model. Additionally, external forces, like wind, Coriolis, friction etc. cause the instantaneous sea surface to deviate from the mentioned equipotential surface in practice. The attraction from planets and its influence on the sea level known as ocean and earth tides can be modelled rather well. The remaining deviation, representing ocean currents, is of the order ± 1 m, so relatively small compared to the geoid heights. Hereafter we assume that the ocean currents are either modelled, or neglected. The latter implies that an additional error of the order of the size of the currents is present.

Under the previously mentioned conditions geoid heights can be obtained from altimetry and in principle one of the main objectives of geodesy seems to be achieved. However, only a discrete set of geoid heights along the typical ground track pattern of the satellite is obtained in ocean areas. The real purpose is to represent the continuous gravity potential worldwide.

Therefore the altimeter data need to be combined with land gravity data. This so called altimetric–gravimetric boundary value problem is not easy to solve, so that a different strategy can be followed. First, create a merged gravity data set at sea indirectly derived from the observed geoid heights and on land directly observed. And subsequently, evaluate Stokes' integral globally by least-squares collocation in a similar way as mentioned in the previous paragraph (cf. Rapp, 1993). Because of the fact that the observations are discrete and gravity is obtained from geoid heights through differentiation the problem is underdetermined and instable. By choosing an appropriate analytical covariance model and inclusion of the measurement noise in the mathematical model one is able to tackle both problems. However, the signal content of the continuous gravity field is band-limited in the sense that only wavelengths can be recovered up to a maximum defined by the sampling pattern or the low pass filter characteristic of the colocation formula.

Apart from the geoid determination two additional advantages of deriving gravity from altimetry should be mentioned. First, gravity from altimetry can directly be compared with observed ship-borne gravity. These comparisons show for several local surveys rms differences of 5–7 mgal (1 mgal = 10^{-5} m/s^2). The accuracy of the local surveys is ±1–1.5 mgal (Rapp, 1985; Haagmans, 1988). These differences are mainly caused by the mentioned band limitation of the altimeter derived gravity since local ship surveys are generally very dense. Nowadays, by combining altimeter data from different satellites one should be able to diminish the difference. One has to keep in mind that altimeter satellites repeatedly measure the whole earth with uniform quality, whence ship expeditions are time consuming, of variable quality, and usually restricted to specific smaller regions.

Secondly, gravity has a direct relation with topography and local or regional density anomalies in the earth's crust. Therefore altimeter derived gravity maps can be used for geophysical or geological interpretation. An example is shown for the Banda Sea near Indonesia in fig. 3. The upper figure represents the continuous gravity field derived from Geosat altimeter data using least-squares collocation, as described in the previous section. The lower figure shows the sea bottom topography for the same area. One can very clearly see the Weber Deep on the right where the depth changes from approximately −1500 m to over −7000 m within ±100 km. The gravity field reflects this deep rather good with a change of ±300 mgal over ±100 km; one of the largest gradients on earth. It seems that the gravity field is rather smooth compared to the sea bottom topography but this is mainly caused by the coarse sampling of the altimeter profiles, ±150 km apart.

Least-squares collocation can be a valuable tool in representing the con-

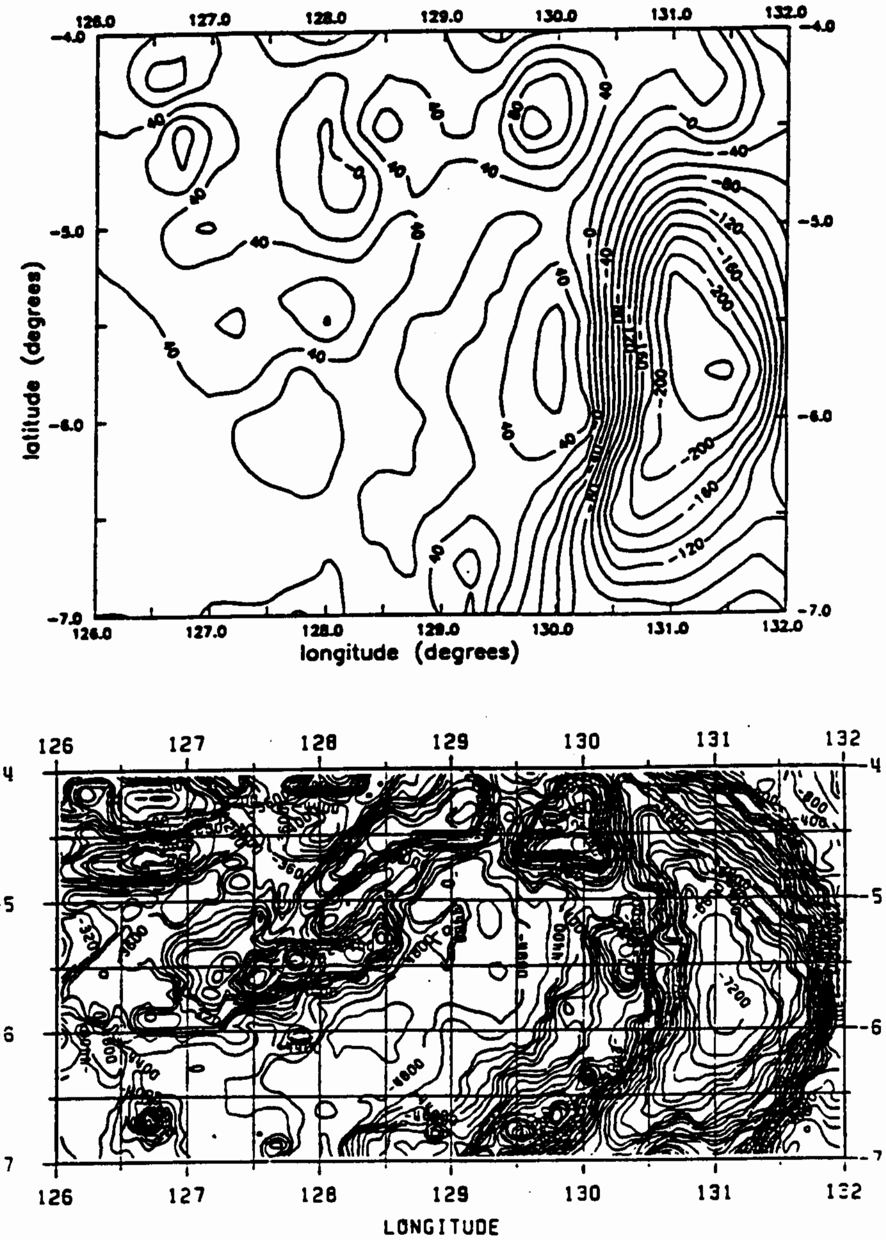

Fig. 3. Upper figure: gravity anomalies obtained from Geosat altimetry in the Banda Sea near Indonesia. Lower figure: sea bottom topography in the same area. On the right is the Weber Deep.

tinuous gravity field in ocean areas from satellite altimeter data. However, an appropriate choice of a covariance model is mandatory, but very difficult. Also the amount of data can be a limiting factor in the use of least-squares collocation as already mentioned in the previous paragraph.

5. SATELLITE TRACKING AND GLOBAL GEOPOTENTIAL MODELLING

In the two foregoing sections least-squares collocation was applied to overcome the inherent underdeterminancy in gravity field estimation problems. For global gravity estimation similar techniques are used, although known under a different name and with different interpretation.

For global gravity field modelling the geopotential coefficients, i.e. the coefficients of a spherical harmonic expansion of the earth's gravitational potential, are taken as parametrisation of the unknown field. Local terrestrial gravity data and satellite orbit information are the primary source of information; especially the latter because of its global nature and the lack of a homogeneous coverage of the earth with gravity data. But the satellite orbits do not provide all the information needed for the solution of the gravity field. The satellite does not 'see' locally, only the total impact of the gravity field on its orbit can be determined by tracking the satellite from ground-based or space-based reference stations. Moreover, the smoothing of the gravity field with increasing altitude, severely hampers the determination of details of the gravity field; see (Reigber, 1989).

The underdetermined problem of the solution of geopotential coefficients from satellite orbit data is found in regularisation of the normal matrix. A (arbitrary) matrix is added to the normal matrix such that it can be inverted. If this matrix contains a-priori covariance information of the unknown potential coefficients this method is identical to collocation. It can also be interpreted as adding zero observations of the unknowns with error (co)variances as given by the additional matrix. These a-priori covariances are obtained from an analytical model, which was estimated from previous solutions of the gravity field. Also it has to be noted that the addition of zero observations introduces a bias into the solution. All geopotential fields determined up to now suffer especially from biases in the weakly determined coefficients of high degree (Xu, 1992). The solution can be optimized by introducing a scaling parameter for the a-priori covariance matrix. Although some coefficients are weakly determined, or are not observable at all, other coefficients are observed very well, those of low degrees (long wavelength) and some at resonance frequencies. From this overdetermined subset, a-posteriori covariances can be computed for the observation noise. As the observation noise can be modelled rather well, the additional parameter can be set such that the a-posteriori variances of the observation equal the a-priori variances.

The quality of a state-of-the-art field, like JGM1-S, can be seen in fig. 4. It shows the ratio between the estimated absolute value of the coefficients and their estimated standard deviation. The figure clearly shows that only up to degree and order 25, apart from some resonant frequencies, significant results could be reached.

When also terrestrial gravity information and other data are added to the satellite orbits, better results can be obtained. Gravity fields up to degree and order 360 are now widely in use in geodesy. But they suffer from the same de-

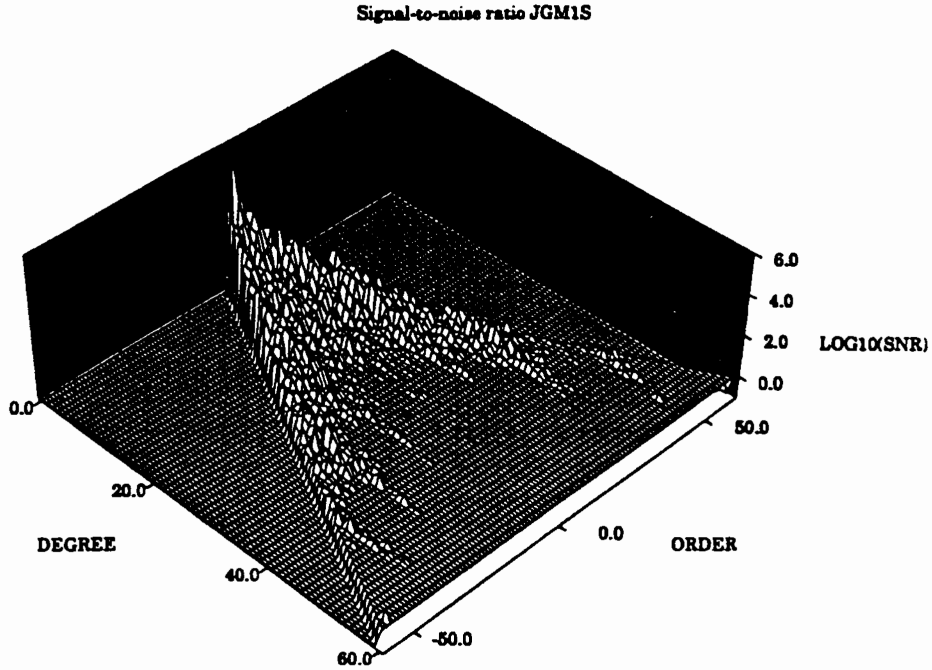

Fig. 4. Log10(SNR) for every coefficient of JGM1-S.

ficiency as the one shown in fig. 4: a strong dependency of the solution on the regularisation matrix. This can only be overcome with a global, homogeneous gravity data set such as can be obtained by the planned ESA satellite gravity gradiometer mission STEP.

6. DISCUSSION

Collocation is a versatile and widely applied tool in physical geodesy, as can be seen from the previous examples. Its strength is that under many conditions reasonable results can be reached. Data of mixed type and quality are combined in an optimal way, provided the covariance information we supply is correct.

This directly leads to the disadvantages and dangers of collocation. Reliable covariance information is not always available. This optimality is not guaranteed anymore, although still reasonably *looking* output is produced. If the function to be estimated has not a zero mean, e.g. for potential coefficients or for local geoid computations with an unsufficient removal of the long-wavelengths, the outcome will be biased.

Furthermore, there are two numerical problems: a very large matrix has to be inverted, its dimension is proportional to the number of observations, which becomes unstable if the data density is too large with respect to the signal content.

REFERENCES

Basic, T. – Untersuchungen zur regionalen Geoidbestimmung mit 'dm' Genauigkeit. Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Nr. 157 (1989).

Haagmans, R.H.N., G.J. Husti, P. Plugers, J.H.M. Smit and G.L. Strang van Hees – NAVGRAV Navigation and Gravimetric Experiment at the North Sea. Netherlands Geodetic Commission, Publication on Geodesy no. 32. Meinema B.V., Delft (1988).

Rapp, R.H. – Detailed Gravity Anomalies and Sea Surface Heights Derived from GEOS-3/Seasat Altimeter Data. Reports of the Department of Geodetic Science and Surveying, no. 365. OSU, Columbus, Ohio (1985).

Rapp, R.H. – Use of Altimeter Data in Estimating Global Gravity Models. In: R. Rummel and F. Sansò (eds.), Satellite Altimetry in Geodesy and Oceanography, Lecture Notes in Earth Sciences 50. Springer-Verlag, Berlin, Heidelberg (1993).

Reigber, Ch. – Gravity field recovery from satellite tracking data. In: F. Sansò and R. Rummel (eds.), Theory of satellite geodesy and gravity field determination. Springer-Verlag, Berlin, 197–234 (1989).

Xu, P. – The value of minimum norm estimation of geopotential fields. Geophys. J. Int. 111, 170–178 (1992).