From Eötvös to mGal

Study Team 2 - Workpackage 3

Scientific Data processing algorithms - the spacewise approach

Final report
29 January 2000

C.C.Tscherning, G.Moreaux
Department of Geophysics, University of Copenhagen

D.Arabelos
Aristotele University of Thessaloniki

A.Albertella, F.Migliaccio, F.Sanso'
DIIAR, Politecnico di Milano

Summary

The software elements able to produce GOCE products using either Least Squares Collocation (LSC) or the Integration approach are described. The software should be made available using a distribution system and the quality standards developed in the EU project MANICORAL.

A number of problems remain unsolved, and require further investigations and software development. Here are identified some of the numerical problems. Since the number of individual observations will be of the order 10^7, normal point or area values will have to be formed, reducing the dataset to 160000 (corresponding to estimating the spherical harmonic coefficients up to degree 400). The normal values may be used in an integration procedure if the normal values are located on the same sphere.

If methods like LSC are used, a system of equations with a full coefficient matrix must be solved. Here we expect that the use of preconditioning based on approximate kernels which are zero outside a certain radius will make it possible to handle such systems. The procedure has been tested on datasets with up to 13000 values distributed in a limited region. The results indicate that the method may permit us to handle much larger systems.

The Slepian approach for coefficient determination is an alternative to the usual harmonic analysis performed using integration when there are gaps in the data, e.g. on the Poles. The approach has been implemented and tested on gravity gradient data at 250 km altitude. The test verified that the Slepian approach is superior to the harmonic approach.

3.1 Introduction.

In the space-wise approach two methods have been identified as being able to produce the main results expected from GOCE, e.g. estimates of the spherical harmonic coefficients of the Earths gravity potential. Also other quantities are needed like gravity anomalies and geoid heights at the Earths surface as well as error estimates of all these quantities.

The two methods are the method of Least Squares Collocation (LSC) and the integration method. The integration method integrates blocks of area mean values (of e.g. T_{zz}) at a fixed altitude. Such mean values may be produced from SGG and SST data in and around the area using LSC. We will denote these values normal mean values. If we want to estimate spherical harmonic coefficients up to e.g. degree 400, at least 401^2 (approximately 160000) normal values have to be formed, including values estimated on the polar caps from ground or airborne gravity surveys.

LSC may also be used globally, but not using simultaneously the more than 10^7 values which probably will be collected. Here along-track normal mean values may be formed in the same process where the normal mean values are formed. Also for LSC 160000 values are needed for a global solution, which should in a satisfactory manner represent all the information collected by GOCE.

This requires a full system of equations with minimally 160000 unknowns to be solved. If Cholesky decomposition were to be used, storage for the upper triangular part of the matrix had to be available, i.e. about 25 GB storage must be available. This is no problem today but if we would try to extract a solution of even higher resolution, problems would arise.

Fortunately the use of the conjugate gradient method with sparse preconditioners seems to be a solution to the problem (Moreaux, 2000). These preconditioners are based on kernels which becomes zero at a certain distance, e.g. 1 degree. This will result in a banded matrix structure with bandwidth between 10 and 20. This means that 20 * 160000 entries have to be in storage. The entries of the full matrix are also needed, but they may be computed "on the fly".

In the integration approach, has well as in LSC, the lack of data on the polar caps creates an error (if ground data is not used). This error may be minimalized in the integration approach by using Slepian functions instead of usual spherical harmonics. The Slepian approach has been implemented and tested.

In the following we will describe the software elements needed to carry out the two types of computations, and some first principle for the software management are sketched.

3.2. Software elements.

The following programs must be available:

(1) A program implementing Least Squares Collocation, and using GOCE-data as well as ground or airborne data and being able to estimate systematic errors (biases, tilts). It must be

able to "predict" normal area or track-segment mean values of GOCE-observables, spherical harmonic coefficients and error estimates. The program should be able to handle up to 160000 observations.

Such a program (GEOCOL) is available as a part of the GRAVSOFT package (Tscherning et. al. 1994), and is in its present version estimated to be able to handle about 50000 observations. (This maximum is due to the limited disk capacity of the local computer system). The equations may be solved using the conjugate gradient method with sparse preconditioners. Software to do this is available (Moreaux, 2000), but should be integrated with the GRAVSOFT programs.

(2) A program implementing the integration approach from normal block averages.

The prototype of such a program using either spherical harmonics or Slepian functions is available at DIIAR, see Albertella and Migliaccio (1994), Albertella and Sneeuw, (1999).

(3) Programs for gross-error detection implementing the two methods described in WP 4.

The GRAVSOFT programs GEOCOL and GEOGRIDE have the capability to perform area-wise gross-error detection. At DIIAR a prototype program is available for fast along-track error detection. But further development is necessary, see WP 4.

(4) Programs for preprocessing (format conversion, reference frame conversion), and for estimating and modelling local and global parameters for the stochastic models (see WP 2).

The GRAVSOFT programs EMPCOV and COVFIT executes estimation of empirical covariance functions and covariance fitting, respectively.

3.3. Software distribution and standards document.

The Web-based distribution system used in MANICORAL (Tscherning et al., 1998) is well suited for GOCE data reduction. A page is established on the web with references to all documents at the different sites, which are engaged in GOCE data management. On the individual sites a user will find each piece of software fully documented and having test in- and output available for download. All source-texts must be available, and reports on test of the software on various platforms must be available also. Timing estimates should be given.

Details about a proposed software standard are given in an appendix to this report.

Preferably the programs should be written in C or FORTRAN.

3.4. Numerical problem areas.

Obviously the largest problem will be the solution of the system of equations having at least 160000 unknowns. However current studies (cf. Section 3.2) have confirmed that this is possible using the conjugent gradient method with sparse preconditioners. Whether numerical problems (related to the condition number) will occur is uncertain. This should be

investigated in simulation studies.

The reduction of the number of unknown from the total number of observations to a number of unknowns in the range of 160000 requires the formation of normal area or track mean values. This is a very heavy computational task, which must be repeated when an GOCE observation cycle is finished. However the task could be distributed to several computers working in parallel, having access to the same observational data base. The design of this database with fast access will be an important task. Here again experience gained in the MANICORAL project may be used.

Another severe problem is to ensure that the signal to noise ratio is correct. If a homogeneous-isotropic statistical model is used, it is not possible to achieve this for the determination of a global solution. Regional solutions may have a statistical model adapted to the region.

Since many programs take advantage of spherical approximation (the Earth is considered a sphere with radius R_E, and the length of the radius vector, r, is calculated as $r = R_E + h$, where h is the altitude) methods for conversion to the "true" values must be verified. However, the integration method may very well be executed rigorously on a sphere at satellite altitude. For the LSC method there are unsolved problems if ground and satellite data are to be treated simultaneously in a global computation without using spherical approximation.

As described in WP 2, SST observations have to be converted to accelerations based on double differences if the existing software is to be used (Tscherning et al., 1990). It should be possible to work directly on the phase information, thereby avoiding the increase in the noise level of a factor two. It would be worthwhile to make this development, which requires that a proper method is found for the numerically integration of covariances along track.

3.5. Implementation of recent developments.

The prediction of spherical harmonic coefficients using LSC has been implemented in a new version of the GEOCOL program. The implementation required the development of new routines for the computation of covariances between the coefficients and the various GOCE data types (Tscherning and Arabelos, 2000).

The prediction can be made using different computational strategies. A fast method for the simultaneous estimation of many coefficients was implemented, but it makes it impossible compute simultaneously the error estimates. However, for this purpose another strategy has been found, see section 3.6.

As mentioned above the use of preconditioning has been implemented in a separate program, which will have to be merged with other programs like GEOCOL. It is not strictly necessary that the preconditioning procedures becomes integrated in the GEOCOL program. An "off-line" solution of the normal-equations is a possibility which should be considered.

The prediction of coefficients have been implemented and used in WP 2 for initial studies using limited datasets. At the moment calculations are done in spherical approximation.

The use of Slepian functions has been tested (Albertella and Sneeuw, 1999) on data using an expansion to degree 180 to create T_{zz} values on a 0.5 degree grid at 250 km altitude. The test showed the superiority of the Slepian approach to the Harmonic approach.

3.6. Software testing.

Tests using data set with 13000 unknowns have been made using the preconditioning method, (Moreaux, 2000).

Controlled tests of coefficient prediction and error-estimation have been made for coefficients of degree up to 180 and using a limited number (5000) of observations. The tests reviled a number of problems related to the estimation of errors of the spherical harmonic coefficients. The most important conclusion is that the quantities to be estimated have to be improvements of the coefficients, and not the coefficients proper. This also enables error estimates to be obtained by computing a series of solutions each obtained by perturbing the observations. Details are given in Tscherning (2000), which is an appendix to this report.

The test of the software using the Slepian approach has been made as described above.

Originally it was the intention to test the developed software on simulated data produced by the ESA project "GOCE end-to-end performance analysis". It has unfortunately not been possible to carry out this optional task, but the tests should be made as soon as possible. Furthermore it was planned to carry out tests on datasets of realistic size. Also this has not been possible due to delays in software development, but it is also important that this is done in the near future.

3.7. Conclusion.

Software elements and software developments needed in order to extract results from GOCE using the space-wise approach have been identified. Promising tests have been made which show that the approach is able to extract the full scientific information from the GOCE data

New integration procedures have been developed and should be explored in order to minimize the numerical integration error. In the LSC method (prediction of coefficients, sparse preconditioners) have been tested on medium size datasets. These developments should be tested on realistic datasets produced by the GOCE end-to-end simulator.

References.

Albertella, A. and F.Migliaccio: Simulated observations of a gradiometric mission: test on the data. Atti del XIII Convengo Nazionale GNGTS, Roma, 1994.

Albertella, A. and N.Sneeuw: The analysis of gradiometric data with Slepian functions. Presented EGS General Assembly, The Hague, 1999.

Moreaux, G.: Harmonic Spherical Splines with Locally Supported Kernels. Proceedings IAG General Assembly, Birmingham, 1999 (in print).

Moreaux, G.: Some preconditioners of harmonic spherical spline problems. Department of Geophysics, University of Copenhagen, (to be submitted), 2000.

Tscherning, C.C.: Computation of spherical harmonic coefficients and their error estimates using Least Squares Collocation. Department of Geophysics, University of Copenhagen, 2000.

Tscherning, C.C. and D.Arabelos: A study of the correlations between spherical harmonic coefficients and point related quantities. Abstract submitted to EGS General Assembly, Nice, April 2000.

Tscherning, C.C., R.Forsberg and M.Vermeer: Methods for regional gravity field modelling from SST and SGG data. Reports of the Finnish Geodetic Institute, No. 90:2, Helsinki, 1990.

Tscherning, C.C., J.Nielsen, D.Arabelos, R.Haagmans, P.Knudsen, F.Sanso' & H.Suenkel: Computer supported Cooperation- the future of the IAG Information service and of cooperation within IAG. Science Services, IAG Scentific Assembly, Rio de Janeiro, Sept. 1997, pp. 47 - 54, IAG 1998.

Proposed software standards.

1. All programs must be written in C or in FORTRAN. If MATLAB is used, it must be verified that the programs run for large datasets.

2. The programs must be well documented with references in the program to the used mathematical or algorithmic source. The source texts must be available.

3. The programs should contain information about author and when the last update was made. A log of changes should be kept describing the major changes.

4. The programs should be tested on several standard platforms, both UNIX (LINUX) and Win98 (or later operating systems). Timing estimates must be given.

5. The programs should accept data in the GRAVSOFT standard format, i.e. data organized as <identifier> <latitude in decimal degrees> <longitude in decimal degrees> <ellipsoidal or orthometric height in m, negative for underwater gravity> <data # 1> ... <data # n>

6. All units must be SI, or as specified (mGal, EU).

7. The format for the description of the attitude of the satellite should be standardized (roll, tilt, pitch, in decimal degrees).

8. Test data and test output should be available which tests all branches of the programs, and so that the output can be compared with an authorized version.