Functional and stochastic modelling of satellite gravity data

Functional and stochastic modelling of satellite gravity data

Jasper van Loon

Publications on Geodesy 67

NCG Nederlandse Commissie voor Geodesie Netherlands Geodetic Commission

Delft, October 2008

Functional and stochastic modelling of satellite gravity data
Jasper van Loon
Publications on Geodesy 67
ISBN: 978 90 6132 307 5
ISSN 0165 1706

Published by: NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission, Delft, The Netherlands
Printed by: Optima Grafische Communicatie, Optima Graphic Communication, Rotterdam, The Netherlands
Cover illustration: Jasper van Loon

NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission
P.O. Box 5058, 2600 GB Delft, The Netherlands
T: +31 (0)15 278 28 19
F: +31 (0)15 278 17 75
E: info@ncg.knaw.nl
W: www.ncg.knaw.nl

The NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission is part of the Royal Netherlands Academy of Arts and Sciences (KNAW).

Acknowledgements

This thesis is the result of a 5-year PhD research at the Delft University of Technology. The support of many colleagues and former colleagues needs to mentioned. First of all, my supervisor Jürgen Kusche. His open communication, enthusiasm and expertise have been of great value to this work and to my own personal development as a researcher. Moreover, he initiated the software development of the energy balance approach, the variance component estimation and the inversion of the GPS site displacements.

Many thanks go my promotor Roland Klees. His feedback and support, especially since Jürgen Kusche left our group, are highly appreciated. Furthermore, I would like to thank all my colleagues at the Physical and Space Geodesy group. Special thanks to my room mate Bas Alberts for all the fun and fruitful discussions we had during our MSc and PhD periods. I will not forget the conversations with Relly van Wingaarden on Barbie dolls, horses and Chinese climbing walls. Moreover, she has been of great support throughout the years for all the non-scientific tasks. I should also mention Mark-Willem Jansen for the discussions we had on the inversion of GPS and GRACE observations and for all the small talks we had on different subjects. Many thanks go to Rene Reudink for all the technical support, Xianglin Liu for the company during the many hours we worked together in the weekends and evenings, and Shizhuo Liu for all the funny conversations we had throughout the years.

I would like to thank the members of the committee for proofreading this thesis. Special thanks to Michael Sideris for his constructive feedback on editorial issues. The CHAMP satellite data was kindly provided by the IAPG in Munich, the GRACE data was downloaded from the CSR data base and the Swiss GPS/levelling data was provided by Urs Marti from the Federal Office of Topography Swisstopo.

However, this thesis would never have been completed without the support of my family and friends. I will not forget all the nice words I received during the period of illness of my sister Barbara. I am lucky that I had plenty of possibilities to relax; playing volleyball, the funny activities at Scouting Halsteren, and my trips to Africa.

I would like to thank my parents for their endless support, but most importantly for creating a place and atmosphere I consider home. The last words of thanks go to my sister Barbara. Her strength in difficult periods has been an example for many, including me.

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Motivation . . . 4
- 1.3 Outline . . . 6

2 Estimation of the Earth's gravity field **9**
- 2.1 The Earth's gravity field . . . 9
  - 2.1.1 Newton's Law of Universal Gravitation . . . 9
  - 2.1.2 Expansion into spherical harmonics . . . 11
  - 2.1.3 Gravity and geoid . . . 13
  - 2.1.4 The gravimetric geoid . . . 15
  - 2.1.5 GPS / levelling . . . 16
- 2.2 The time-variable gravity field . . . 18
  - 2.2.1 The inverse problem . . . 18
  - 2.2.2 Elastic loading of the Earth . . . 20
- 2.3 Gravity field determination from satellite data . . . 23
  - 2.3.1 Satellite tracking . . . 23
  - 2.3.2 Satellite altimetry . . . 25
  - 2.3.3 Satellite gradiometry . . . 25
- 2.4 The determination of data weights in satellite gravity field modelling . . . 26
  - 2.4.1 The early years of satellite gravity field modelling (1958-1975) . . . 27
  - 2.4.2 External calibration (1975-1988) . . . 28
  - 2.4.3 Internal calibration (1988-2000) . . . 29
  - 2.4.4 Dedicated satellite missions (2000-) . . . 31
- 2.5 Summary . . . 33

3 Augmentation of the functional model **35**
- 3.1 The Least-Squares principle . . . 35
  - 3.1.1 The Least-Squares estimator of the vector of unknowns . . . 35
  - 3.1.2 The overall model test . . . 37
- 3.2 Global and local parameters . . . 38
  - 3.2.1 The estimation of global and local parameters . . . 38
  - 3.2.2 Test of significance for local parameters . . . 40

3.3 The addition of prior information . . . 43
3.3.1 Augmentation of the functional model . . . 43
3.3.2 Test for consistency with prior information . . . 43
3.4 Softly unbiased estimation . . . 44
3.5 Summary . . . 47

4 Stochastic model validation **49**
4.1 The stochastic model . . . 50
4.2 Minimum Norm Quadratic Unbiased Estimator (MINQUE) . . . 51
4.3 Properties of the MINQUE under normality . . . 55
4.3.1 Best Invariant Quadratic Unbiased Estimator (BIQUE) . . . 56
4.3.2 Least-Squares Variance Component Estimation (LSVCE) . . . 56
4.3.3 Restricted Maximum Likelihood Estimation (REML) . . . 58
4.3.4 Other properties . . . 61
4.4 Iterative Restricted Maximum Likelihood Estimator (IREML) . . . 63
4.5 Uncorrelated observation groups . . . 65
4.5.1 Groups with multiple variance components . . . 65
4.5.2 Groups with only one variance component . . . 70
4.6 Alternative weighting algorithms . . . 72
4.6.1 Helmert's Variance Component Estimation . . . 72
4.6.2 The Bayesian estimate and the MAP estimate . . . 74
4.6.3 Lerch's subset solution method . . . 75
4.6.4 Generalized Cross-Validation (GCV) . . . 78
4.6.5 (Iterative) Maximum Likelihood Estimator (MLE and IMLE) . . . 78
4.6.6 External calibration . . . 80
4.7 Summary . . . 80

5 Monte Carlo implementation **83**
5.1 Stochastic Trace Estimation (STE) . . . 83
5.2 Monte Carlo MINQUE (MC-MINQUE) . . . 85
5.2.1 Groups with multiple variance components . . . 85
5.2.2 Groups with only one variance component . . . 88
5.3 Monte Carlo IREML (MC-IREML) . . . 92
5.3.1 Groups with multiple variance components . . . 92
5.3.2 Groups with only one variance component . . . 94
5.4 Monte Carlo Lerch (MC-Lerch) . . . 94
5.5 Summary . . . 96

6 Outlier detection and robust estimation **97**
6.1 Test statistics in outlier detection . . . 98
6.1.1 The w-test statistic . . . 98
6.1.2 Pope's test statistic . . . 99
6.1.3 The μ-test statistic . . . 100
6.1.4 Treatment of the outliers . . . 100
6.2 Detection of outliers in correlated observations . . . 100

6.3 Robust estimation . . . 101
6.3.1 M-estimation of the vector of unknowns . . . 103
6.3.2 Choice of the distribution . . . 104
6.3.3 Cost Function Estimation (CFE) . . . 107
6.3.4 M-estimation of the variance components . . . 110
6.4 Summary . . . 113

7 Application 1: CHAMP satellite gravity data 115
7.1 The energy balance approach for gravity recovery . . . 115
7.2 Test setup 1: One data set (232 days) . . . 117
7.3 Re-weighting of the data sets . . . 119
7.4 Treatment of outliers . . . 126
7.5 Test setup 2: Two year data set . . . 129
7.6 Combination with a prior gravity model . . . 132
7.6.1 The EGM96 global gravity model . . . 132
7.6.2 Stochastic model validation (first iteration) . . . 132
7.6.3 Augmentation of the functional model . . . 136
7.6.4 Stochastic model validation (second iteration) . . . 137
7.7 Summary and outlook . . . 139

8 Application 2: Joint inversion of global GPS time-series with GRACE gravity models 141
8.1 The functional model of the joint inversion . . . 142
8.2 Inversion of weekly GPS displacements . . . 143
8.2.1 Description of the GPS data . . . 143
8.2.2 Constraining the solution . . . 144
8.2.3 Weekly solutions . . . 146
8.2.4 Outlier treatment . . . 147
8.2.5 Covariance information . . . 149
8.2.6 Monthly solutions . . . 151
8.2.7 Modified regularization matrix . . . 151
8.3 Inversion of monthly GRACE gravity models with weekly GPS displacements . 154
8.3.1 Description of the GRACE data . . . 154
8.3.2 GPS/GRACE combination . . . 157
8.4 Evaluation of the local parameters . . . 158
8.4.1 Test for significance . . . 160
8.4.2 Motion of the geocentre . . . 160
8.5 Summary and outlook . . . 160

9 Application 3: Temporal aliasing of hydrological signals in a simulated GRACE recovery 165
9.1 Temporal aliasing of signals . . . 165
9.2 Test setup . . . 166
9.2.1 Hydrological model . . . 167
9.2.2 Simulation of the noise . . . 167
9.3 Results . . . 168

9.3.1 The temporal aliasing effect . . . 168
9.3.2 The addition of random noise . . . 170
9.3.3 The combination of noise and hydrological signal . . . 173
9.4 Summary and outlook . . . 175

10 Application 4: The computation of a height reference surface in Switzerland 177
10.1 Test setup . . . 177
10.1.1 The data sets used . . . 179
10.1.2 The four-parameter corrector surface . . . 179
10.2 Validation of the stochastic model . . . 180
10.2.1 Variance Component Estimation . . . 180
10.2.2 Diagonalized variance-covariance matrices . . . 182
10.3 Treatment of outliers . . . 182
10.3.1 Conventional 3-sigma rule . . . 182
10.3.2 Robust M-estimation . . . 182
10.4 Other parametrizations of the corrector surface . . . 185
10.5 Summary and outlook . . . 188

11 Conclusions and recommendations 191
11.1 Conclusions . . . 191
11.2 Recommendations . . . 194

References 197

A Series expansion into spherical harmonics 217

B Matrix algebra and matrix analysis 219

C Some standard distributions 221

Summary 223

Samenvatting 227

Curriculum Vitae 231

Symbols

Relation symbols

$A := B$	:	A is defined as B.
$A \approx B$	:	A approximates B.
$A \propto B$	:	A is proportional to B.
$A \sim B$	:	A is distributed as B.
$A \neq B$	:	A is not equal to B.

Operators

$E\{A\}$	:	Expectation of A.
$C\{\underline{a}, \underline{b}\}$	:	Correlation between $\underline{a}$ and $\underline{b}$.
$D\{\underline{a}\}$	:	Dispersion of $\underline{a}$.
$MSE\{\underline{a}\}$	:	Mean Square Error of $\underline{a}$.
$p(a)$	:	Probability of a.
$p(a\|b)$	:	Probability of a under the assumption that b occurs.
$\mathcal{R}(A)$	:	Range of A.
$\mathcal{N}(A)$	:	Null space of A.
tr(A)	:	Trace of A.
det$A$	:	Determinant of A.
vec$\{A\}$	:	vec operator on A.
vh$\{A\}$	:	vh operator on A.
$A \otimes B$	:	Kronecker product between A and B.
A^T	:	Transpose of A.
A^{-1}	:	Inverse of A.
$A^{\perp}$	:	Orthogonal complement of A.
div$A$	:	Divergence of A.
∇A	:	Gradient of A.
ΔA	:	Laplace operator on A.

Scalars ($\in \mathbb{N}$)

l	:	Degree of a spherical harmonic function.
L	:	Truncation degree of spherical harmonic model.
m	:	Order of a spherical harmonic function.
m	:	Number of observations.
n	:	Number of unknown parameters.
p	:	Number of observation groups.
q	:	Number of variance components to be estimated.
r	:	Redundancy.
δ_{ij}	:	Kronecker delta.

Scalars ($\in \mathbb{R}$)

a	:	Radius of the spherical shell, in which the surface mass changes occur [m].
$\bar{C}_{lm}$, $\Delta\bar{C}_{lm}$	:	Spherical harmonic coefficient.
C	:	Arbitrary integration constant.
f	:	Flattening of the Earth's ellipsoid.
G	:	Gravitational constant = $6.6742 \cdot 10^{-11}$ $[\mathrm{Nm^2kg^{-2}}]$.
h'_l	:	Height Load Love Number for degree l.
h	:	Ellipsoidal height $[m]$.
H	:	Orthometric height in point P $[\mathrm{m}]$.
k	:	Threshold in the outlier detection algorithms.
k'_l	:	Gravitational Load Love Number for degree l.
l_{PQ}	:	Distance between points P and Q $[\mathrm{m}]$.
l'_l	:	Lateral Load Love Number for degree l.
m_Q	:	Mass of Q $[\mathrm{kg}]$.
M	:	Mass of the Earth, scale factor of the geopotential $[\mathrm{kg}]$.
N	:	Geoid height $[\mathrm{m}]$.
$P_{lm}(\cos\vartheta)$	:	Legendre function of degree l and order m.
$\bar{P}_{lm}(\cos\vartheta)$	:	4π-normalized Legendre function of degree l and order m.
Q_l	:	Molodenskii truncation coefficients.
r	:	Radius $[\mathrm{m}]$.
r_i	:	Redundancy number of group i.
R	:	Radius of the Earth, scale factor of the geopotential $[\mathrm{m}]$.
s	:	Helmert scaling parameter.
s_i	:	Denominator of Lerch's method.
$\bar{S}_{lm}$, $\Delta\bar{S}_{lm}$	:	Spherical harmonic coefficient.
T	:	Disturbance potential $[\mathrm{m^2/s^2}]$.

Scalars ($\in \mathbb{R}$) (continued)

U	:	Reference or normal potential $[\mathrm{m}^2/\mathrm{s}^2]$.
v_i	:	Calibration or variance factor of group i.
V	:	Gravitational potential $[\mathrm{m}^2/\mathrm{s}^2]$.
$\hat{w}_i$	:	w-test statistic of observation i.
W	:	Gravity potential $[\mathrm{m}^2/\mathrm{s}^2]$.
$W(\alpha)$	:	Weighting kernel in an isotropic smoothing.
$W(\vartheta, \lambda, \vartheta', \lambda')$	:	Weighting kernel in a non-isotropic smoothing.
W_l	:	Coefficients of the weighting kernel $W(\alpha)$.
δh_i	:	Discrepancy in GPS-levelling data $[\mathrm{m}]$.
$\delta\Omega$	:	Test statistic, difference between two residual square sums.
$\Delta e(\vartheta, \lambda)$	:	Displacement in East direction of (ϑ, λ).
$\Delta g(P)$	:	Gravity anomaly in point P $[\mathrm{m}/\mathrm{s}^2]$.
$\Delta h(\vartheta, \lambda)$	:	Height displacement in (ϑ, λ).
$\Delta n(\vartheta, \lambda)$	:	Displacement in North direction of (ϑ, λ).
$\Delta\sigma(\vartheta, \lambda)$	:	Surface mass in (ϑ, λ).
$\Delta\Psi(\vartheta, \lambda)$	:	Poloidal component of the lateral displacement in (ϑ, λ).
$\gamma(P)$	:	Normal gravity in point P $[\mathrm{m}/\mathrm{s}^2]$.
$\eta(\vartheta, \lambda)$	:	Equivalent water height in (ϑ, λ).
ϑ	:	Co-latitude $[\mathrm{rad}]$.
λ	:	Longitude $[\mathrm{rad}]$.
λ	:	Non-centrality parameter in hypothesis testing.
$\hat{\mu}_i$	:	μ-test statistic of observation i.
ν_i	:	Robust weight of observation i in the M-estimation.
$\rho(Q)$	:	Density in point Q $[\mathrm{kg}/\mathrm{m}^3]$.
$\rho(\mu_i)$	:	Cost function value of observation i.
σ^2	:	Overall model test.
σ_i^2	:	Variance component.
σ_{ij}^2	:	Covariance component.
$\hat{\tau}_i$	:	Pope's test statistic of observation i.
Ψ_{PQ}	:	Spherical distance between points P and Q $[\mathrm{rad}]$.
$\Psi(M)$	:	Lagrange function, with M as free parameter.
$\Psi(\hat{\mu}_i)$	:	Influence value of observation i.
ω_e	:	rotational speed of the Earth $[\mathrm{rad}/\mathrm{s}]$.
Ω	:	Residual square sum.

Vectors

$\underline{a}_P$	:	Gravitational acceleration vector in point P [$\mathrm{m/s^2}$].
$\underline{b}$	:	Vector of Helmert transformation parameters.
$\underline{e}$	:	Vector of residuals.
$\underline{e}_{PQ}$	:	Unit vector between points P and Q.
$\underline{F}_P$	:	Gravitational force vector in point P [N].
$\underline{h}$	:	Right-hand-side vector of system of normal equations.
$\underline{h}_{[i]}$	:	Right-hand-side vector, omitting observation group i.
$\underline{h}_s$	:	R.h.s. vector of the global parameters in a reduced system of normal equations.
$\underline{l}$	:	Vector to define the linear combination of variance components to be estimated.
$\underline{q}$	:	Right-hand-side vector in the estimation of Helmert VCE.
$\underline{\hat{q}}_i$	:	Output vector of Monte Carlo VCE.
$\underline{r}_P$	:	Vector from origin to point P.
$\underline{t}$	:	Auxiliary vector used in the derivation of the REML and IREML estimators.
$\underline{u}$	:	R.h.s. vector of the system of normal equations to solve the variance components.
$\underline{v}$	:	Vector of misclosures.
$\underline{x}$	:	Vector of unknown parameters.
$\underline{y}$	:	Vector of observations.
$\underline{z}$	:	Random vector.
$\underline{\gamma}$	:	Vector of variance components.
$\underline{\delta}$	:	Vector of discrepancies.
$\underline{\epsilon}$	:	Vector of Helmert rotation parameters.
$\underline{\epsilon}_i$	:	Vector of elementary errors of group i.
$\underline{\Psi}_i$	:	Reduced global right-hand-side vector of observation group i.

Matrices

A	:	Design matrix.
C_i	:	Cofactor matrix of group i.
D	:	Duplication matrix, i.e. $D \cdot \text{vh}\{M\} = \text{vec}\{M\}$.
D_i	:	Result of factorization $W_i = D_i D_i^T$.
F_i	:	Auxiliary matrix to derive Helmert VCE.
G_i	:	Result of factorization $C_i = G_i G_i^T$.
G_{ij}	:	Sub-matrix of Q_y^{-1}.
H	:	Defines the linear relationship between the prior information and $\underline{x}$.
H	:	Normal matrix in the estimation of the variance components using REML-nr.
H	:	Normal matrix in the estimation of the variance components using Helmert VCE.
I	:	Identity matrix.
$\mathfrak{I}_{\underline{\gamma}}$	:	Fisher's information matrix.
J_i	:	Auxiliary matrix to simplify the equations for disjunctive observation groups.
L	:	Auxiliary matrix to derive the IREML.
M	:	Symmetric matrix, to be optimized in a Quadratic Variance Component Estimation.
N	:	Normal matrix.
$N_{[i]}$	:	Normal matrix, omitting observation group i.
$\bar{N}_{ss}$	:	Normal matrix of the global parameters in a reduced system of normal equations.
O	:	Matrix, of which the elements have value zero.
P_A	:	Projection matrix, which projects a vector onto the column space $\mathcal{R}(A)$.
$P_A^{\perp}$	:	Projection matrix, which projects a vector along $\mathcal{R}(A)$.
Q_i	:	Cofactor matrix of group i.
Q_t	:	Variance-covariance matrix of $\underline{t}$.
$Q_{\hat{x}}$	:	Variance-covariance matrix of $\underline{\hat{x}}$.
Q_y	:	Variance-covariance matrix of the observations.
$Q_{\hat{\gamma}}$	:	Variance-covariance matrix of the estimated variance components.
R	:	Rao's or redundancy matrix.
S	:	Normal matrix of system of normal equations to solve for the variance components.
T_i	:	Auxiliary matrix to derive Helmert's VCE.
U_i	:	Linear relationship between group i of elementary errors and $\underline{e}$.
W_i	:	Weight matrix of observation group i.
Υ_i	:	Reduced normal matrix of the global parameters, belonging to observation group i.

1

Introduction

1.1 Background

The gravitational field of the Earth shows large fluctuations, due to the heterogeneous mass distribution of the Earth, its oblateness, the topography and the dynamical processes within the Earth. Moreover, a time-varying gravitational signal can be measured, caused by mass redistributions on the surface of the Earth (due to, e.g., hydrology) and within the atmosphere and the oceans. The gravitational field describes the physical shape of the Earth, perturbates the orbit of a satellite, and gives extra information on the dynamics of the Earth, the oceans, and its cryosphere; see, e.g., Rummel et al. [2002]. An important surface is the *geoid*, i.e. the equipotential surface, which best fits the mean sea surface at rest. The difference of the geoid with respect to a reference ellipsoid varies between -100 metres (Indian Ocean) and 80 metres (Iceland); see figure 1.1.

According to Stokes's theorem [Heiskanen and Moritz, 1967], it is not possible to determine the mass distribution of the Earth uniquely from its external gravity field. However, the large majority of the mass redistributions over geologically short (daily to inter-annual) timescales, takes place in a thin layer near the Earth's surface [Chao, 2005]. Therefore, these temporal gravity changes can almost be directly inverted to land hydrology, ocean dynamics and changes in the cryosphere. An exception is the effect of the post-glacial rebound (PGR) on the time-varying gravity signal. Temporal gravity changes are often expressed in equivalent water heights, which take amplitudes up to several decimeters. Seasonal water storage in large river basins, like the Amazon and the Zambezi, are clearly visible in data of the Gravity Recovery and Climate Experiment (GRACE) mission [Tapley et al., 2004a]. Moreover, recent studies, e.g., Velicogna and Wahr [2006] and Luthcke et al. [2006], show a significant decrease in polar ice mass (Antarctica and Greenland), based on GRACE gravity measurements.

After the removal of the influence of topography on the Earth's static gravity field, the remaining static field will give us information on the mass distribution of the Earth's

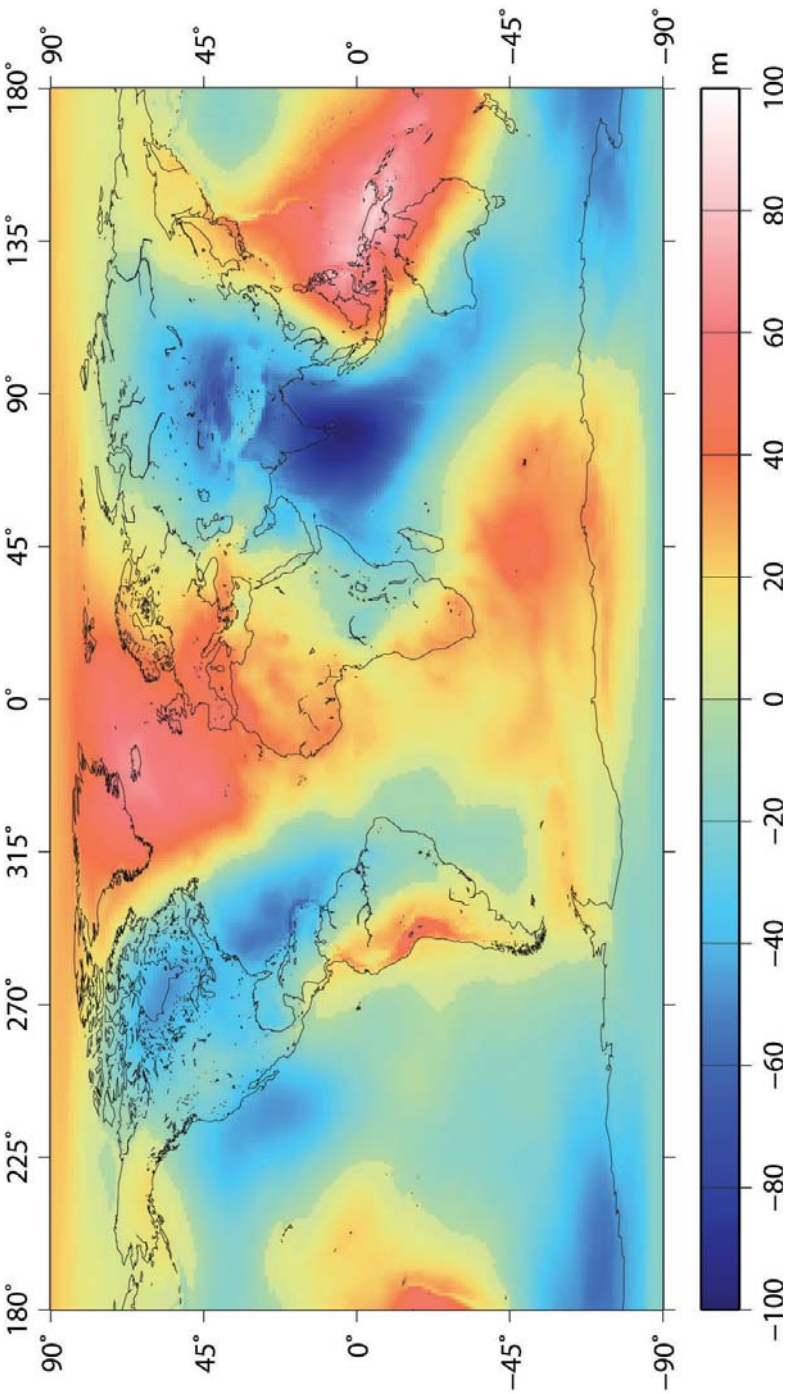


Fig. 1.1: The EIGEN-GL04C geoid model, up to degree and order 360, i.e., with a spatial resolution of 111 kilometres (shortest wavelength) [Förste et al., 2008].

lithosphere and mantle. Together with seismic tomography, deformation measurements (e.g InSAR), magnetometry and laboratory research on mantle material, the gravitational field is one of the few sources that can give insight into the properties and structure of the lithosphere and its interaction with the (upper) mantle. A summary of these dynamical processes is shown in figure 1.2.

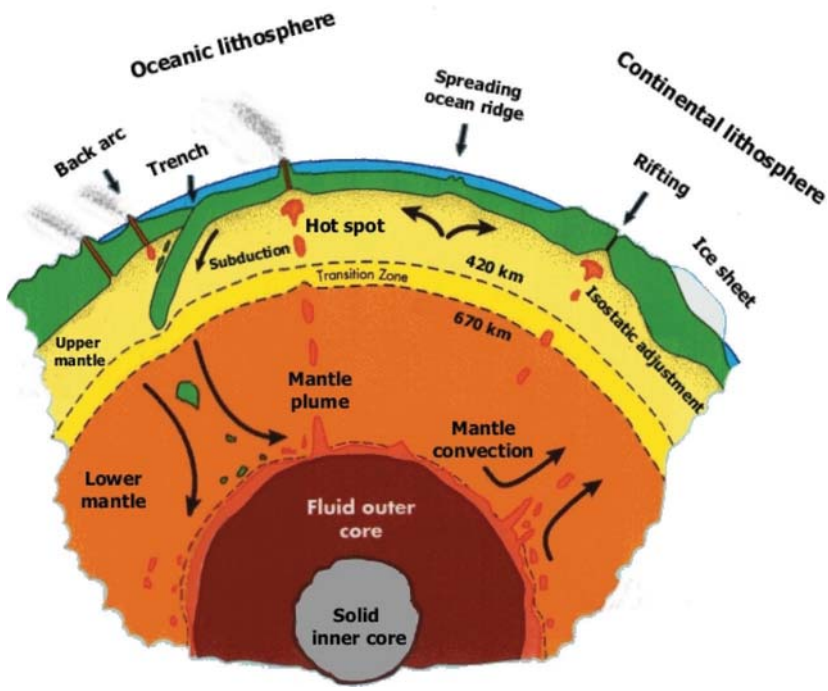


Fig. 1.2: The dynamical processes within the Earth's mantle and lithosphere [BGI].

Research on global climate changes, including sea level rise, is largely dependent on the knowledge of the ocean surface [Le Grand and Minster, 1999]. If one wants to monitor ocean currents, together with mass and heat transportation within the oceans, one needs to know what the influence of the heterogeneous mass distribution is on the ocean surface. This influence, expressed in the geoid, should be subtracted from altimetric measurements of the sea surface to obtain the mean dynamic topography; see ,e.g., Vossepoel [2007]. Pietrzak et al. [2007] have used satellite altimetry observations, together with an accurate geoid model, to monitor the destructive tsunami wave of December 2004 in the Indian Ocean. A study on the effect of the preceding earthquake on the gravity field of the Earth is discussed in Han and Simons [2008].

The gravitational field is essential for the height determination of points on the Earth; in absolute terms, because the geoid is used as a global reference level, and in relative terms, because gravity defines the direction in which water flows and in this way defines which points are higher than other points. An accurate geoid model can, at least in theory, be used to transform purely geometric heights, measured by GPS, into physical heights.

The motion of a satellite orbiting the Earth is mainly caused by the gravitational field, together with non-conservative forces, like atmospheric drag and solar radiation pressure. In dynamic orbit determination, one makes use of a global gravity model, together with positioning techniques like GPS and satellite laser ranging (SLR). The inverse approach is to determine a global (homogeneously accurate) gravity model from orbit information of several satellites. This approach has been used since the launch of the first satellites, see, e.g., Buchar [1958], in which the oblateness of the Earth could be derived from an alternative data source, compared to the terrestrial and marine observations used so far. Satellite data has been the key data source for global gravity models, especially in recent years, with the launch of two dedicated gravity missions, the Challenging Mini-satellite Payload for Geophysical Research and Applications (CHAMP) mission and the GRACE mission, and with the upcoming Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission, to be launched in September 2008. These new satellite observations have already improved our knowledge about the static and time-varying gravity field of the Earth and have given new insights into the large-scale processes in the interior of the Earth and the smaller (time-varying) changes on its (sub-)surface.

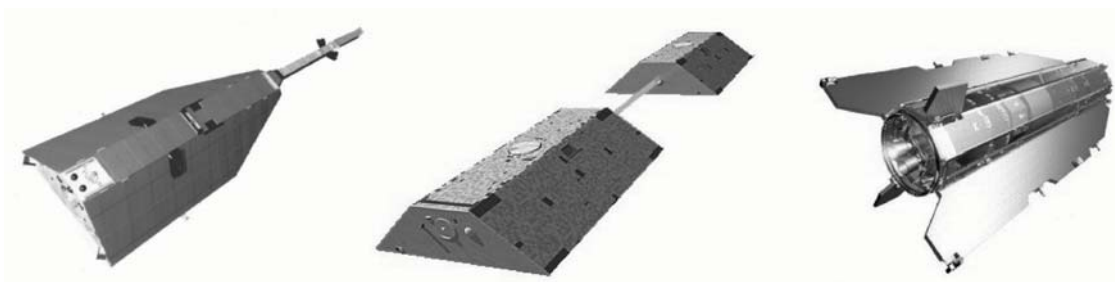


Fig. 1.3: CHAMP, GRACE, and GOCE satellite missions.

1.2 Motivation

It is common practice in gravity modelling and precise orbit determination (POD) to combine different data sets in a least-squares sense. The purpose is to fill in the gaps in the space domain (e.g., the polar gap of satellite data), strengthen parts of the frequency domain or extend the observed time interval. Combinations are possible using data from different satellite missions (e.g., CHAMP plus GRACE), from different arcs or periods within one mission, from different observation types (e.g., satellite data and terrestrial data) or from a combination of new data with prior information about the unknown gravity field parameters.

Before combining the different data sets, one should pay attention to a proper modelling of the functional and stochastic model and one should either remove outliers or reduce their impact on the least-squares solution. The functional model expresses the

(linearized) relationship between the observations and the vector of unknown parameters. This vector could include the parameters of the global gravity model, expressed in spherical harmonics, as well as local parameters per data set, often applied to absorb systematic effects. The data sets can be inconsistent with each other, in the space or the frequency domain, and these inconsistencies should be statistically tested before implemented into the functional model.

If the functional model is well-defined, one can use the residuals between the observations and the estimated model to estimate the stochastic model. In this way, we can re-weight different data sets and obtain a better estimate of the stochastic properties of the data sets. These properties may vary considerably in time, due to instrument conditions, the changing GPS constellation, the applied preprocessing and other factors. The stochastic model is used in the weighted least-squares estimation (WLSE) and the detection of outliers. Moreover, it propagates into the quality description of the estimated parameters and derived quantities.

To improve the stochastic model, we write the variance-covariance (VC) matrix of the observations as a linear combination of different (cofactor) matrices and estimate these linear coefficients (i.e., the *variance components*). In the early seventies, C.R. Rao [Rao, 1971a, 1973] developed the Minimum Norm Quadratic Unbiased Estimator (MINQUE), which, under normality, has minimum variance [Sjöberg, 1983], converges to the solution of (restricted) maximum likelihood [Koch, 1986] and can be derived in a least-squares sense [Teunissen, 1988].

Variance component estimation (VCE) techniques have been used in many geodetic applications, e.g., in the combination of ellipsoidal, orthometric and geoid heights to derive a corrector surface [Fotopoulos, 2003, 2005], in SLR to give weights to the different SLR stations [Sahin et al., 1992], in very long baseline interferometry (VLBI) to re-weight the data sets [Lucas and Dillinger, 1998], in a simulated GOCE orbit [Koch and Kusche, 2002], and in the regularization of ill-posed problems; see, e.g., Koch and Kusche [2002] and Xu et al. [2006].

However, data weighting in global gravity field modelling has often been based on experience or trial and error. In the 1990's, several groups had used an alternative weighting method suggested by F.J. Lerch [Lerch et al., 1988], which is an internal calibration technique based on subset solutions. Since the launch of the dedicated satellite missions, not much effort has been placed into the validation of the stochastic model. In this field of research, the geodetic community has made a step back in history, due to the elaborate processing of large amounts of data. Groups are using an equal weighting of the data sets or use the stochastic properties of the observations defined by the industry, which are in general not accurate and oftenly too optimistic.

In this thesis, we try to promote the use of VCE techniques and provide tools to use any least-squares software package to solve for the variance components. To do this, we randomize the input vector and use Monte Carlo (MC) techniques to estimate the variance components. The resulting method is simple and each researcher can use his own software package to retrieve the variance components and consequently obtain a better estimate of the stochastic model.

Outliers are observations that do not match the suggested functional and stochastic model. They will deteriorate the least-squares solution of the vector of unknowns and of the estimated variance components. One often uses hypothesis testing to detect possible outliers and consequently removes them. The threshold defines which observations are good and which are bad. An alternative to this is not to remove the spurious observations, but to down-weight those observations, which have a large test statistic. This so-called M-estimation technique has been used in several geoscience applications, such as SLR data processing [Yang et al., 1999], magnetic Ørsted satellite data processing [Olsen, 2002], relative GPS positioning [Chang and Guo, 2005], and GPS network adjustment [Yang et al., 2005].

With this robust technique, we actually assume a different distribution of the observations than the normal distribution. In this way, a higher probability is given to outliers, reducing the effect of such an observation in the least-squares estimation. We suggest to use the huge redundancy in the observations to estimate the probability density function and use this new function to detect and down-weight possible outliers.

1.3 Outline

We will start (chapter 2) with some theoretical and historical background on satellite gravity field modelling. In section 2.1, the theory of universal gravitation, spherical harmonics, data combination and GPS-levelling is explained, which will be used throughout this thesis, especially in the second part of the thesis in which we deal with some satellite gravimetry applications. Special attention is paid on the temporal gravity field of the Earth (section 2.2). Several techniques are explained in section 2.3 to retrieve gravity-related information from satellite observations, including SLR, altimetry, satellite-to-satellite tracking (SST), and satellite gravity gradiometry (SGG). A historical overview of the data weighting problem in the estimation of global gravity models is the topic of the remaining part of chapter 2.

Augmentation of the functional model to include inconsistencies between different data sets, systematic errors or other local parameters, is the subject of chapter 3. Special focus is on the combination of the new satellite data with prior information. The chapter starts with the principles of least-squares and hypothesis testing (section 3.1). The augmentation of extra (local) parameters to the functional model to account for e.g. systematic effects and the test for significance of these parameters are the subject of section 3.2. Section 3.3 deals with the addition of prior information and the corresponding tests for unbiasedness between the two data sets. The chapter ends with the softly unbiased estimation, which weakens the constraints for unbiasedness in the combination with prior information.

Chapter 4 is the key chapter of this thesis. It starts (section 4.1) with the parametrization of the stochastic model in terms of a linear combination of cofactor matrices, in which the variance components are the linear coefficients. The most well-known estimator of these variance components is the MINQUE. Derivation of this estimator

is done in section 4.2. It is shown in section 4.3 that under normality this estimator has minimum variance, (restricted) maximum likelihood and can be derived in a weighted least-squares sense. The derivation of the Least-Squares Variance Component Estimation (LSVCE) was done in a recent PhD study within the DEOS Department [Amiri-Simkooei, 2007]. A much faster estimator, which converges to the same results as MINQUE (under normality), but is biased in the iterations, is the Iterative Restricted Maximum Likelihood Estimator (IREML); see section 4.4. The equations can be simplified if we assume several disjunctive observation groups. Section 4.5 deals with groups of multiple variance components and groups with only one variance component to be estimated per observation group. Several estimators of the variance components have been used in the past in satellite gravity field modelling, which, in general, do not converge to the MINQUE-like estimators. These alternative weighting algorithms include Helmert VCE, Bayesian estimation, Lerch's subset solution method, Maximum Likelihood Estimation (MLE) and external calibration techniques. These methods will be dealt with in section 4.6.

The use of Monte Carlo techniques to speed up the computations is the topic of chapter 5. The chapter provides useful tools to use existing least-squares software in the computation of the variance components. Both MINQUE and IREML estimators as well as Lerch's subset solution technique are dealt with.

Chapter 6 discusses the detection and treatment of outliers. We start in section 6.1 with the derivation of the different test statistics to detect the outliers. A discussion of outlier detection in the presence of coloured noise is the subject of section 6.2. The conventional approach is to remove the outliers which fails the hypothesis testing. In section 6.3, we suggest an alternative approach, in which we down-weight the outliers. The theory of M-estimation is explained and a modification of Fellner's approach [Fellner, 1986] is suggested to robustify the VCE. Moreover, the estimation of the probability density function from the obtained residuals is discussed.

The first application, to be discussed in chapter 7, is the global gravity field modelling using CHAMP pseudo-observations from the energy balance approach (EBA). In this application, several data weighting algorithms are compared. Comparisons are made on the solutions (at convergence), the computation time, and the quality assessment of the results. The estimation of the stochastic model is combined with a treatment of the outliers. Conventional data snooping is compared with the robust M-estimation technique and the estimation of the probability density function. Before combining the satellite-only solution with a prior gravity model, several inconsistency vectors are tested and, if necessary, implemented in the functional model. The stochastic model of the prior information is written as a linear combination of several cofactor matrices and the variance components are estimated.

The second application (chapter 8) is a joint inversion of weekly GPS site displacements with monthly GRACE solutions to derive surface mass changes, expressed in equivalent water heights [Blewitt et al., 2001]. With such a combination, we can use the advantages of both data sources to strengthen the combined solution. Weekly estimates of the motion of the geocentre could be derived as a by-product.

The effect of temporal aliasing is addressed in chapter 9. We try to quantify the effect and see if it is possible to reduce the distortions in the estimation of the temporal gravity field. For this, we simulated the annual signal of a hydrological model along simulated GRACE orbits.

The fourth application is the estimation of a local *height reference surface* of Switzerland, which can be used in the determination of orthometric heights from ellipsoidal GPS heights (chapter 10). Several heterogeneous data sets are involved in this estimation, each with its own stochastic properties.

The thesis ends with conclusions and recommendations (chapter 11).

2

Estimation of the Earth's gravity field

In this chapter we will show how to model the Earth's gravity field with the available gravity-related observations. First, we will consider some basics of potential theory and introduce the definitions of geoid and gravity. Different techniques to derive gravity information from satellite observations are explained. The chapter ends with a historical overview of the data weighting problem in the determination of global gravity models from satellite data.

2.1 The Earth's gravity field

2.1.1 Newton's Law of Universal Gravitation

Two objects in space will be attracted to each other by the force of *gravitation*. Sir Isaac Newton was the first who showed that the motion of the Moon around the Earth can be explained by the same force as an object falling to the Earth. He introduced the Law of Universal Gravitation, which states the following [Newton, 1687]:

> *Every object in the Universe attracts every other object with a force directed along the line of centres for the two objects that is proportional to the product of their masses and inversely proportional to the square of the distance between the two objects.*

This law is based on experiments. The source of gravitation, necessary for an analytical derivation of this law, is still under discussion. Newton's law means that point mass Q, with mass m_Q $[kg]$ will generate a gravitational force $\underline{F}$ on point mass P with mass m_P

and a distance l_{PQ} $[m]$ to point Q, which is equal to

$$\underline{F}(P) = -G\frac{m_P m_Q}{l_{PQ}^2} \cdot \underline{e}_{PQ}, \tag{2.1}$$

where G is the gravitational constant, estimated at $6.6742 \cdot 10^{-11}\ Nm^2kg^{-2}$ [Mohr and Taylor, 2005], and $\underline{e}_{PQ}$ the unit vector pointing from P to Q. The acceleration of the point mass m_P due to the gravitational force of m_Q can easily be computed:

$$\underline{a}(P) = \frac{\underline{F}(P)}{m_P} = -G\frac{m_Q}{l_{PQ}^2} \cdot \underline{e}_{PQ}. \tag{2.2}$$

Hence, the gravitational acceleration is independent of the mass of the attracted body. This acceleration vector can be written as the gradient of a scalar function V, which is called the *gravitational potential*, i.e.,

$$\underline{a}(P) = \nabla V(P) \tag{2.3}$$

where this scalar function is defined by

$$V(P) := G\frac{m_Q}{l_{PQ}} + C, \tag{2.4}$$

with C an arbitrary constant, which we will define as $C := 0$ from now on to ensure $V = 0$ for $l_{PQ} \to \infty$. Note that the potential itself cannot be measured, only potential differences. The focus of gravity field determination is on the estimation of this scalar function V. The potential induced by multiple point masses is equal to the sum of the potentials of each mass separately (superposition principle):

$$V(P) = G\sum_{i=1}^{N} \frac{1}{l_{PQ_i}} m_i \tag{2.5}$$

For an arbitrary body Σ_Q, composed of infinitively small point masses $\mathrm{d}m_Q$, we can state that $\mathrm{d}m_Q = \rho(Q)\mathrm{d}\Sigma_Q$ and

$$V(P) = G\int_{\Sigma_Q} \frac{\rho(Q)}{l_{PQ}} \mathrm{d}\Sigma_Q, \tag{2.6}$$

with $\rho(Q)$ the mass density $[\mathrm{kg/m^3}]$ at point Q. The Laplace operator Δ, applied to the scalar function V, is defined as

$$\Delta V = \mathrm{div}\nabla V, \tag{2.7}$$

which is expressed in Cartesian coordinates as

$$\Delta V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \tag{2.8}$$

The gravitational potential V satisfies *Poisson's equation* [Heiskanen and Moritz, 1967], i.e.,

$$\Delta V(P) = -4\pi G\rho(P). \tag{2.9}$$

Outside the sources, where $\rho(P) = 0$, this equation becomes

$$\Delta V(P) = 0. \tag{2.10}$$

This is called *Laplace's equation*. A function, which fulfills Laplace's equation, is a harmonic function. Stokes's theorem states [Heiskanen and Moritz, 1967] that

> *A function V, harmonic outside Σ_Q, is uniquely determined by its values on the boundary. On the other hand, however, there are infinitely many mass distributions inside Σ_Q, which have a given V as exterior potential.*

So, if we can describe the harmonic function V at a surface enclosing the masses, we can compute this harmonic function at any point in space. However, it will not be possible to determine the mass distribution from only external gravity information.

2.1.2 Expansion into spherical harmonics

To model the external gravitational potential of a body (e.g., the Earth), one needs to know the harmonic function on an enclosing surface of the body of mass. This information is, however, only known at some discrete points on this surface. It is, therefore, necessary to approximate the potential by a function with a limited number of parameters, which can be estimated by the discrete points. Any harmonic function can be represented by a series of spherical harmonics [Heiskanen and Moritz, 1967]:

$$V(\vartheta, \lambda, r) = \sum_{l=0}^{\infty} r^{-(l+1)} \sum_{m=0}^{l} (a_{lm} \cos m\lambda + b_{lm} \sin m\lambda) P_{lm}(\cos\vartheta) \tag{2.11}$$

where $\cos m\lambda \cdot P_{lm}(\cos\vartheta)$ and $\sin m\lambda \cdot P_{lm}(\cos\vartheta)$ are the base functions, called *spherical harmonics*. For more information about the Legendre functions $P_{lm}(\cos\vartheta)$ and the orthogonality properties, we refer to appendix A. When determining the gravitational field of the Earth, it is common practice to use the series expansion

$$V(\vartheta, \lambda, r) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} (\bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda) \bar{P}_{lm}(\cos\vartheta), \tag{2.12}$$

in which the Legendre functions $P_{lm}(\cos\vartheta)$ are replaced by the 4π-normalized Legendre functions $\bar{P}_{lm}(\cos\vartheta)$ (appendix A) and the coefficients are rescaled to better suit the estimation of the Earth's gravity field numerically. The parameter M is an estimate of the total mass of the Earth including the atmosphere and R an estimate of the radius of the Earth, although one should see both GM and R mainly as scale factors of the

geopotential model [Smith, 1998]. As low-orbiting satellites are not capable of improving the estimate of GM, one often set $\bar{C}_{00} := 1$. Another constraint can be put into this series if the origin of the coordinate system is located at the centre of mass. Then $\bar{C}_{10} = \bar{C}_{11} = \bar{S}_{11} = 0$. The number of coefficients are still unlimited. We can bound this number by truncating the model to a degree L. The error, which is introduced when truncating the model, is called the *omission error* or the *truncation error*. The influence of errors in the potential coefficients is called the *commission error*. The first global models were truncated to a degree 8. Nowadays it is possible to derive global models up to degree and order 2160 with the use of terrestrial data; see, e.g., the recently released EGM08 global gravity model [Pavlis et al., 2008]. A higher truncation degree will result in a model with a higher resolution. However, the highest power of the gravitational potential can be found in the lowest degrees. *Parseval's theorem* states

$$\frac{1}{4\pi R^2}\int_{\sigma_R} V^2 \mathrm{d}\sigma_R = \left(\frac{GM}{R}\right)^2 \sum_{l=0}^{\infty}\sum_{m=0}^{l}\left(\bar{C}_{lm}^2 + \bar{S}_{lm}^2\right), \tag{2.13}$$

in which the left-hand-side represents the mean energy of the gravitational potential V on σ_R, with σ_R the surface of a sphere with radius R. The right-hand-side includes the dimensionless degree-variances σ_l^2:

$$\sigma_l^2 := \sum_{m=0}^{l}\left(\bar{C}_{lm}^2 + \bar{S}_{lm}^2\right). \tag{2.14}$$

Hence, the signal degree variances σ_l^2 represent the contribution of a certain degree to the mean energy of the gravitational potential on σ_R. The dimensionless signal degree variances of the gravitational potential are often modelled by an analytical function called *Kaula's rule* [Kaula, 1966]:

$$\sigma_l^2 \approx \frac{1.6 \cdot 10^{-10}}{l^3}. \tag{2.15}$$

With such an analytical expression, one is able to estimate the omission error. Other models exists, such as the Tscherning / Rapp model [Tscherning and Rapp, 1974]. The mean signal variance per coefficient can easily be derived as

$$\sigma^2(\bar{C}_{lm}, \bar{S}_{lm}) = \frac{\sigma_l^2}{2l+1} \approx \frac{10^{-10}}{l^4}. \tag{2.16}$$

This variance is often used in satellite-only models of the Earth's gravitational field. Discrete measurements at high altitude, and/or a polar gap for non-polar orbits, can cause an ill-posed problem; see, e.g., Tikhonov and Arsenin [1977] and Bouman [1998a]. The high-degree (and low-order) potential coefficients are poorly resolved, attain too much power and have high mutual correlations. A well-known technique to solve this problem, used in most satellite-only solutions and even in some combined solutions,

is to add zero-valued pseudo-observations of the potential coefficients to the system of observation equations. Kaula's rule is used to assign noise variances to these pseudo-observations (Eq. (2.16)). This can be viewed as a *Tikhonov regularization*. It decreases the correlations between the coefficients and stabilizes the ill-posed problem. A drawback of the constraints is the biasedness of the solution and the decrease in power in the higher degrees [Xu, 1992]. More on this *modified least-squares* approach can be found in Moritz [1980] as a special case of least-squares *collocation*.

2.1.3 Gravity and geoid

If one measures gravity with a gravimeter, one does not only measure the gravitational force of the Earth's mass. Other forces are also present and need to be either modelled or measured and subsequently removed from the gravimeter observations. The largest disturbing force is the centrifugal force of the rotating Earth. The centrifugal acceleration is perpendicular to the axis of rotation and maximum at the equator. The sum of gravitation and the centrifugal force is called *gravity*. The gravity potential, therefore, reads

$$W(\vartheta,\lambda,r) = \frac{GM}{R}\sum_{l=0}^{\infty}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}(\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda)\bar{P}_{lm}(\cos\vartheta) + \frac{1}{2}\omega_e^2 r^2\sin^2\vartheta. \tag{2.17}$$

As the rotation of the Earth ω_e is well-known and does not change much in time, the centrifugal force can be modelled quite easily and subsequently be subtracted from the observations.

Gravimeters located on moving platforms, such as ships or airplanes, will measure a Coriolis acceleration when moving in longitudinal direction (again due to the rotating Earth). Gravity measurements taken on board of moving objects need to be corrected for this effect (*Eötvös-correction*), which can be done if the velocity of the object is sufficiently known.

The Earth forms a two-body system with the Moon, in which it is attracted by the gravitational field of the Moon and in which it rotates around their common barycentre. The Earth-Moon system forms a similar two-body system with the Sun. The difference between the gravitational attractions and centrifugal forces due to these systems is called the *direct tide* and can be modelled quite easily. An indirect effect of the tidal forces are the *ocean tides*. Recent models show quite accurate estimates of these ocean tides, which have to be removed from the gravity data.

Other contributions, such as atmospheric and hydrological variations, are more difficult to model as they continuously change in time.

Every mass will be attracted by its surrounding masses and will be accelerated along the gradient of the total potential. Ocean water will, therefore, flow according to this rule

until the sea surface has reached an equilibrium state. Although the sea will never be at rest due to currents and winds, we may consider a particular equipotential surface, the *geoid*, as a good approximation of the sea surface (after reduction of the tidal effects).

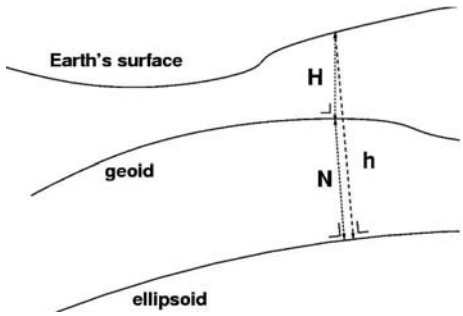


Fig. 2.1: The geoid height N, the elevation H above the geoid and the ellipsoidal height h.

To linearize the functional relationship between gravity measurements and the gravity potential, we need a good approximation of this geoid. This reference (or *normal*) potential is generated by a rotating ellipsoid with its surface being an equipotential surface. It can be defined by only four parameters: the semi-major axis of the ellipsoid (a), the flattening of the ellipsoid (f), the potential value (U_0) at its surface and the rotation parameter (ω_e). The values of this reference potential can be computed with

$$U(\vartheta,\lambda,r) = \frac{GM^U}{R^U}\sum_{l=0,2,4,\ldots}^{\infty}\left(\frac{R}{r}\right)^{l+1}\bar{C}^U_{l0}\cdot\bar{P}_{l0}(\cos\vartheta) + \frac{1}{2}\omega_e^2 r^2\sin^2\vartheta. \tag{2.18}$$

The difference between W and U is called the *disturbance potential* T:

$$T(\vartheta,\lambda,r) = \frac{GM}{R}\sum_{l=2}^{\infty}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}(\Delta\bar{C}_{lm}\cos m\lambda + \Delta\bar{S}_{lm}\sin m\lambda)\bar{P}_{lm}(\cos\vartheta), \tag{2.19}$$

where $\Delta\bar{C}_{lm}$ and $\Delta\bar{S}_{lm}$ are the differences between the coefficients of W and U. We will assume $R = R^U$ and $M = M^U$ and furthermore neglect differences between the potential coefficients for degrees 0 and 1. The geoid heights, i.e. the difference between the geoid and the reference ellipsoid, can be linked to the disturbance potential by ***Bruns' equation***:

$$N(\vartheta_P,\lambda_P) = \frac{T(P_0) - \Delta W_0}{\gamma(P_e)}, \tag{2.20}$$

with P_0 the point of the geoid with geodetic coordinates (ϑ_P,λ_P) and P_e the point on the ellipsoid with the same geodetic coordinates. The quantity $\gamma(P_e)$ is the *normal* gravity at point P_e, i.e., the gravity *computed* from the reference ellipsoid. The potential difference between the geoid and the ellipsoid, ΔW_0, will be assumed to be zero, as we have defined $\Delta\bar{C}_{00} := 0$.

2.1.4 The gravimetric geoid

In this section, we will show how to compute geoid heights from gravity observations using Stokes's approach. We will make use of *gravity anomalies*, i.e.,

$$\Delta g(P_0) := g(P_0) - \gamma(P_e). \tag{2.21}$$

As the gravity measurements are, in general, not performed on the geoid, they need to be reduced to this geoid by a *free-air* reduction; see, e.g., Li and Götze [2001], i.e.

$$g(P_0) \approx g(P) + c^{FA} \cdot H_P \quad ; \quad c^{FA} = 0.3086 \,\mathrm{mGal/m}, \tag{2.22}$$

with $1\,\mathrm{mGal}$ equal to $1 \cdot 10^{-5}\,\mathrm{m/s^2}$ and H_P the *orthometric* height of point P. The gravity anomalies can be linked to the disturbance potential with the *fundamental equation of physical geodesy*:

$$\Delta g(P_0) \approx -\left(\frac{2}{R}T(P_0) + \frac{\partial T}{\partial r}(P_0)\right). \tag{2.23}$$

The geoid heights can be computed from the gravity anomalies by *Stokes's formula* [Heiskanen and Moritz, 1967]:

$$N(\vartheta_P, \lambda_P) = \frac{R}{4\pi\gamma} \int_\sigma St(\Psi_{PQ}) \Delta g_Q \mathrm{d}\sigma_Q, \tag{2.24}$$

with Ψ_{PQ} the angular distance between points P and Q and the Stokes's function defined by

$$St(\Psi_{PQ}) := \sum_{l=2}^{\infty} \frac{2l+1}{l-1} P_l(\cos\Psi_{PQ}). \tag{2.25}$$

The gravity anomalies are assumed to lie on the ellipsoid, although they are computed on the geoid. This approximation is valid as $\dfrac{\partial \Delta g}{\partial r}$ is rather small. As the gravity anomalies will only be known at discrete points on the geoid, one will, in many practical cases, use mean gravity anomalies $\Delta\bar{g}_i$ instead [De Min, 1996],

$$N(\vartheta_P, \lambda_P) \approx \frac{R}{\gamma} \sum_i^{n_i} St(\Psi_{PQ_i}) \Delta\bar{g}_i \frac{\Delta s_i}{4\pi}, \tag{2.26}$$

where the gravity anomalies are averaged over small surface elements Δs_i.

We have seen that we can calculate the geoid heights with the use of *global* potential coefficients (Bruns's equation), or with the use of *local* gravity anomalies (Stokes's formula). As the global satellite gravity models are truncated at a degree L_s, one will introduce a spectral truncation error. Surface gravity data at discrete points contain the full spectrum of the gravity field, but lack a global coverage. However, as they are typically averaged within a block to form mean gravity anomalies, they are as well truncated to a

certain degree L_m. A combined solution will use global potential coefficients to ensure global coverage and surface gravity data to improve the resolution within a certain area.

A commonly used method to combine the global coefficients with the local gravity data is the *remove-restore* technique [Rapp and Rummel, 1975]. With this technique, the low-degree part of the spectrum is computed directly from the spherical harmonic coefficients. The contribution of this part on the gravity anomalies is subtracted from the local gravity data. The residual observations are then used (in Stokes's equation) to improve the geoid heights in a specific area.

An alternative is suggested by Molodenskii et al. [1962]. The local gravity data is used to compute a first approximation of the geoid heights in that specific area. The contribution of the outer region (where there is no local gravity data) is computed using the *Molodenskii truncation coefficients* $Q_l(\Psi_c)$:

$$\Delta N(\vartheta_P, \lambda_P) = \frac{GM}{R\gamma} \sum_{l=2}^{L_s} Q_l(\Psi_c) \sum_{m=0}^{l} (\Delta \bar{C}_{lm} \cos m\lambda_P + \Delta \bar{S}_{lm} \sin m\lambda_P) \bar{P}_{lm}(\cos \vartheta_P), \tag{2.27}$$

of which the coefficients Q_l can be computed recursively as was shown by Shepperd [1982] and Paul [1983]. The contribution ΔN is then added to the initial geoid heights.

2.1.5 GPS / levelling

We will now use the relation between the ellipsoidal height h, e.g., measured with GPS, the geoid height N and the orthometric height H:

$$\delta h_i = h_i - N_i - H_i \quad ; \quad E\{\delta h_i\} = 0. \tag{2.28}$$

Theoretically, if the discrepancies δh_i are close to zero, one could use the combination of ellipsoidal heights h_i and the geoid height N_i to estimate the orthometric heights H_i, and in this way avoid the costly and time-consuming levelling operations [Schwarz et al., 1987]. However, the discrepancy δh_i can, in general, not be neglected. It is caused by several error sources and is a sum of systematic and stochastic errors. More on this subject can be found in, e.g., Featherstone [1998], Kotsakis and Sideris [1999], Fotopoulos [2003] and Prutkin and Klees [2008]. We will now discuss the largest contributions to $\delta h_i \neq 0$.

Errors in the gravimetric geoid

The gravimetric geoid is in general estimated using several different data sources, each with its own stochastic properties. The long-wavelength errors, together with the local (surface gravity) errors, propagate into this gravimetric solution. Moreover, the theory of the determination of the gravimetric geoid makes use of some assumptions and approximations.

Errors in the levelling (orthometric) heights

Apart from the systematic errors (e.g., calibration, atmospheric refractions) and the stochastic errors in the levelling observations, the orthometric heights will also encounter some extra error sources due to the adjustment of the observations to estimate the orthometric heights. One of this error sources is the orthometric correction, which cannot be solved exactly, as one would need continuous gravity measurements along the plumb line. A commonly used approximation is to assume a constant density and use the Poincaré-Prey gravity gradient in the orthometric correction formula to derive Helmert orthometric heights. Another source of error are the constraints needed to correct the levelling observations, as one only measures orthometric height differences between two points and no absolute heights. One could constraint the solution with tide gauges as zero constraints, introducing errors due to sea surface topography and long-period tidal effects, or one could use levelling markers, which are in general not stable and should not be used as fixed constraints.

Errors in the ellipsoidal (GPS) observations

Ellipsoidal heights can be obtained from different systems, such as VLBI, SLR, Doppler measurements (DORIS) and GNSS systems, like GLONASS, GPS and the future Galileo system. The GPS system is the most commonly used system to derive ellipsoidal heights. The measurements my have systematic and stochastic errors, of the order of a few centimetres.

Other error contributions

One should be cautious that one measures the same quantities with all three observation types, both in time domain as well as in space domain. Time differences can cause systematic effects, e.g., due to land subsidence or geodynamic processes. If one does not measure at the same location, e.g., one cannot measure the levelling marker with GPS and the correction is not accurate enough, this will propagate directly into the discrepancy δh_i. Another possible error source is the difference in reference field (ellipsoid) subtracted from the gravimetric geoid and the GPS measurements or the use of a different tide system. A different choice in ellipsoid can be overcome by a five parameter transformation surface; see, e.g., Heiskanen and Moritz [1967] and Kotsakis and Sideris [1999]. Some smaller contributions include the difference in angle between the ellipsoid normal and the plumb line at the geoid (see, e.g., figure 2.1) and the curvature of the plumb line in contrast to the geometrical ellipsoidal heights.

The height reference surface

The discrepancy δh_i is in general too large to simply compute orthometric heights from GPS and geoid heights; see, e.g., Kotsakis and Sideris [1999] and Prutkin and Klees [2008]. A surveyor's approach is to estimate a 'corrector surface' and to add this to the

gravimetric geoid to obtain a transformation surface for the GPS measurements. We will call this the *height reference surface*. The corrector surface, added to the gravimetric geoid, could be e.g., a (tilted) plane, a surface obtained by least-squares collocation [Moritz, 1980], multiquadric interpolation [Hardy, 1971] or the tensioned spline algorithm [Smith and Wessel, 1990]. A division of an area into several subareas to estimate several corrector surfaces is proposed in Jiang and Duquenne [1996]. In Featherstone [1998], it is even questioned if we really need a gravimetric geoid. The alternative would be to estimate a transformation surface from a large number of GPS-levelling reference points.

Despite all possible errors mentioned before, GPS-levelling data can be used as a quality assessment of *global* gravity models, as the data is still much more accurate than any global model; see, e.g., EGM96 [Lemoine et al., 1998] and EGM08 [Pavlis et al., 2008].

2.2 The time-variable gravity field

Knowledge on the time-varying gravity field (e.g., by using data from the GRACE mission) gives us extra insight into the dynamical processes within a small layer near the Earth's surface. Although most of these variations are periodic (e.g., hydrological signal from rain seasons), recent results show secular trends, such as a clear decrease in the mass of the polar ice sheets. A combination of GRACE gravity data with site displacements, measured by GPS, will be discussed in chapter 8. This section will focus on the theoretical background of the temporal gravity signal.

2.2.1 The inverse problem

According to Stokes' theorem, it is not possible to determine the density distribution of a body of mass uniquely from its external gravity field. However, the majority of the mass redistributions occurring within short (daily to inter-annual) timescales take place near the Earth's surface; see, e.g., Chao [2005]. We can, therefore, associate the time-variable gravity field to surface mass redistributions within a thin spherical shell σ_a near the surface ($r = a$). The gravity potential generated by these surface masses $\Delta\sigma_Q$ can be evaluated as

$$\begin{aligned}
\Delta V(P) &= G \cdot \iiint_{\Sigma_Q} \frac{\Delta\rho_Q}{l_{PQ}} \mathrm{d}\Sigma_Q \\
&= G \cdot a^2 \cdot \iint_{S_Q} \frac{1}{l_{PQ}} \int_{\sigma_a} \Delta\rho_Q \mathrm{d}r \mathrm{d}S_Q \quad \text{(where } S_Q \text{ is the unit sphere)} \\
&= G \cdot a^2 \cdot \iint_{S_Q} \Delta\sigma_Q \frac{1}{l_{PQ}} \mathrm{d}S_Q \quad \text{(where } \Delta\sigma_Q = \int_{\sigma_a} \Delta\rho_Q \mathrm{d}r\text{).}
\end{aligned} \tag{2.29}$$

If we expand the inverse of the distance between the points P and Q into a series of spherical harmonics [Heiskanen and Moritz, 1967], i.e.,

$$\begin{aligned}\frac{1}{l_{PQ}} = \; & \frac{1}{r_Q}\sum_{l=0}^{\infty}\frac{1}{2l+1}\left(\frac{r_Q}{r_P}\right)^{l+1}\cdot \\ & \sum_{m=0}^{l}\bar{P}_{lm}(\cos\vartheta_P)\bar{P}_{lm}(\cos\vartheta_Q)(\cos m\lambda_P\cos m\lambda_Q+\sin m\lambda_P\sin m\lambda_Q),\end{aligned} \tag{2.30}$$

if $r_P \geq r_Q$, the potential of a spherical shell ($r_Q = a$) can be re-written as

$$\begin{aligned}V(P) \; = \; & G\cdot a\cdot\iint_{S_Q}\Delta\sigma_Q\sum_{l=0}^{\infty}\frac{1}{2l+1}\left(\frac{a}{r}\right)^{l+1}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\vartheta_P)\bar{P}_{lm}(\cos\vartheta_Q)\cdot \\ & (\cos m\lambda_P\cos m\lambda_Q+\sin m\lambda_P\sin m\lambda_Q)\mathrm{d}S_Q \\ = \; & G\cdot a\cdot\sum_{l=0}^{\infty}\frac{1}{2l+1}\left(\frac{a}{r}\right)^{l+1}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\vartheta_P)\cdot \\ & \Big[\cos m\lambda_P\cdot\iint_{S_Q}\bar{P}_{lm}(\cos\vartheta_Q)\cos m\lambda_Q\Delta\sigma_Q\mathrm{d}S_Q+ \\ & \sin m\lambda_P\cdot\iint_{S_Q}\bar{P}_{lm}(\cos\vartheta_Q)\sin m\lambda_Q\Delta\sigma_Q\mathrm{d}S_Q\Big]\end{aligned} \tag{2.31}$$

The surface mass densities $\Delta\sigma_Q$ [kg/m^2] along the spherical shell σ_a can be expanded into a series of spherical harmonics:

$$\Delta\sigma(\vartheta,\lambda) = a\rho_w\sum_{l=0}^{\infty}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda+\Delta\bar{S}^{\sigma}_{lm}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta) \tag{2.32}$$

with ρ_w the average density of water (set to 1025 kg/m^3) and $\Delta\bar{C}^{\sigma}_{lm}$ and $\Delta\bar{S}^{\sigma}_{lm}$ the dimensionless coefficients, defined as

$$\left\{\begin{array}{c}\Delta\bar{C}^{\sigma}_{lm}\\ \Delta\bar{S}^{\sigma}_{lm}\end{array}\right\} := \frac{1}{4\pi a\rho_w}\iint_{S_Q}\bar{P}_{lm}(\cos\vartheta_Q)\left\{\begin{array}{c}\cos m\lambda_Q\\ \sin m\lambda_Q\end{array}\right\}\Delta\sigma_Q\mathrm{d}S_Q. \tag{2.33}$$

The choice to use the density of water ρ_w in the series of spherical harmonics is not only to make the coefficients dimensionless. The largest part of the mass redistributions is due to the transportation of water around the globe. An alternative is, therefore, to express the mass redistributions in *equivalent water heights* η:

$$\eta(\vartheta,\lambda) = \frac{\Delta\sigma(\vartheta),\lambda}{\rho_w} = a\sum_{l=0}^{\infty}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda+\Delta\bar{S}^{\sigma}_{lm}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta). \tag{2.34}$$

Combining Eq. (2.31) and Eq. (2.33) yields

$$\Delta V(\vartheta,\lambda,r) = \frac{GM}{a}\frac{3\cdot\rho_w}{\rho_e}\sum_{l=0}^{\infty}\frac{1}{2l+1}\left(\frac{a}{r}\right)^{l+1}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda + \Delta\bar{S}^{\sigma}_{lm}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta), \tag{2.35}$$

with the average density of the Earth ρ_e defined as [Blewitt and Clarke, 2003]

$$\rho_e := \frac{M_e}{\frac{4}{3}\pi a^3} \approx 5514\ kg/m^3, \tag{2.36}$$

with M_e the mass of the Earth and a the radius of a sphere with the same volume as the Earth. From now on, we will assume that the scaling parameters of the static field, i.e. M and R, are equal to M_e and a respectively. The time-varying gravity potential can be inverted uniquely into equivalent water heights if the contribution of atmospheric masses can be removed using isobaric geopotential height data [Swenson and Wahr, 2002] and if one assumes that (the majority of) the remaining mass redistributions within short timescales are due to water transportations within a small layer near the surface. Such an inversion is the primary aim of the GRACE satellite mission; see, e.g., Tapley et al. [2004a]. However, it is only possible to detect the *temporal changes* in equivalent water heights, as the static part of the water storage cannot be separated from the solid Earth contribution with the use of satellite gravity data only.

2.2.2 Elastic loading of the Earth

When we assume that all mass redistributions occur in a thin layer at the Earth's surface and if we assume a purely elastic model, we can use the theory of Farrell [1972] to model the Earth's response to the loading masses. In this theory, the *load love numbers* (LLN) only depend on the spherical degree l. The surface load creates an additional potential. The potential due to the surface mass redistributions will, therefore, change to

$$\Delta V(\vartheta,\lambda,r) = \frac{GM}{R}\frac{3\cdot\rho_w}{\rho_e}\sum_{l=0}^{\infty}\frac{1+k'_l}{2l+1}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda + \Delta\bar{S}^{\sigma}_{lm}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta), \tag{2.37}$$

with k'_l the potential LLN of degree l. The LLN and certain combinations of these numbers can be found in table 2.1. The temporal change in the Earth's gravitational

potential, i.e. $\Delta V(\vartheta, \lambda, r)$, can be measured by satellite gravimetry. Therefore, if we compare Eq. (2.37) to

$$\Delta V(\vartheta, \lambda, r) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} [\Delta \bar{C}_{lm}^{V} \cos m\lambda + \Delta \bar{S}_{lm}^{V} \sin m\lambda] \bar{P}_{lm}(\cos \vartheta) \tag{2.38}$$

we can state that the coefficients $\Delta \bar{C}_{lm}^{\sigma}, \Delta \bar{S}_{lm}^{\sigma}$ can directly be retrieved from the gravity potential coefficients $\Delta \bar{C}_{lm}^{V}, \Delta \bar{S}_{lm}^{V}$:

$$\left\{ \begin{array}{c} \Delta \bar{C}_{lm}^{V} \\ \Delta \bar{S}_{lm}^{V} \end{array} \right\} = \frac{3 \cdot \rho_w}{\rho_e} \frac{1 + k_l'}{2l + 1} \cdot \left\{ \begin{array}{c} \Delta \bar{C}_{lm}^{\sigma} \\ \Delta \bar{S}_{lm}^{\sigma} \end{array} \right\} \tag{2.39}$$

Tab. 2.1: Load Love Numbers and combinations to degree 12, from Blewitt and Clarke [2003]

Degree l	$1 + k_l'$	h_l'	l_l'	$\frac{3 \cdot \rho_w}{\rho_e} \frac{1 + k_l'}{2l + 1}$	$\frac{3 \cdot \rho_w}{\rho_e} \frac{h_l'}{2l + 1}$
1	1.021	-0.269	0.134	0.190	-0.050
2	0.693	-1.001	0.030	0.077	-0.112
3	0.805	-1.052	0.074	0.064	-0.084
4	0.868	-1.053	0.062	0.054	-0.065
5	0.897	-1.088	0.049	0.045	-0.055
6	0.911	-1.147	0.041	0.039	-0.049
7	0.918	-1.224	0.037	0.034	-0.046
8	0.925	-1.291	0.034	0.030	-0.042
9	0.928	-1.366	0.032	0.027	-0.040
10	0.932	-1.433	0.030	0.025	-0.038
11	0.934	-1.508	0.030	0.023	-0.037
12	0.936	-1.576	0.029	0.021	-0.035

The Earth itself will also change its shape due to the time-varying gravity potential, which is caused by the hydrological and atmospheric redistributions [Blewitt et al., 2001]. As the shape of the Earth and its deformations can be measured very accurately by GPS using the global network of the International GNSS Service (IGS), this provides us with an extra independent data set; see, e.g., Blewitt et al. [2001] and [Blewitt and Clarke, 2003]. The GPS station displacements [Ferland et al., 2000] can be jointly inverted with the GRACE measurements to estimate the surface mass redistributions. This may improve the estimation of the lower degrees coefficients from GRACE. In Kusche and Schrama [2005], it was shown that in such a joint inversion the contribution of GPS was up to 60% for degrees 2 till 4 and up to 30% for higher degrees coefficients, based upon formal covariances.

The direct change in geoid heights due to surface masses reads

$$\begin{aligned}\Delta N(\vartheta,\lambda) &= R\sum_{l=0}^{\infty}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{V}\cos m\lambda+\Delta\bar{S}_{lm}^{V}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta)\\ &= R\frac{3\rho_w}{\rho_e}\sum_{l=0}^{\infty}\frac{1}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{\sigma}\cos m\lambda+\Delta\bar{S}_{lm}^{\sigma}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta)\end{aligned} \tag{2.40}$$

The Earth will directly react to this by changing the height Δh of the Earth's surface. Using LLN h_l', we obtain:

$$\begin{aligned}\Delta h(\vartheta,\lambda) &= R\frac{3\rho_w}{\rho_e}\sum_{l=0}^{\infty}\frac{h_l'}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{\sigma}\cos m\lambda+\Delta\bar{S}_{lm}^{\sigma}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta)\\ &:= R\sum_{l=0}^{\infty}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{h}\cos m\lambda+\Delta\bar{S}_{lm}^{h}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta)\end{aligned} \tag{2.41}$$

We can, therefore, compute the coefficients $\Delta\bar{C}_{lm}^{\sigma},\Delta\bar{S}_{lm}^{\sigma}$ directly from the coefficients $\Delta\bar{C}_{lm}^{h},\Delta\bar{S}_{lm}^{h}$:

$$\left\{\begin{array}{c}\Delta\bar{C}_{lm}^{h}\\ \Delta\bar{S}_{lm}^{h}\end{array}\right\}=\frac{3\cdot\rho_w}{\rho_e}\frac{h_l'}{2l+1}\cdot\left\{\begin{array}{c}\Delta\bar{C}_{lm}^{\sigma}\\ \Delta\bar{S}_{lm}^{\sigma}\end{array}\right\} \tag{2.42}$$

In this way, we can use GPS height displacements to estimate the surface mass redistributions. The lateral displacement can be retrieved by first expressing its poloidal component

$$\begin{aligned}\Delta\Psi(\vartheta,\lambda) &= R\sum_{l=0}^{\infty}l_l'\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{V}\cos m\lambda+\Delta\bar{S}_{lm}^{V}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta)\\ &= R\frac{3\rho_w}{\rho_e}\sum_{l=0}^{\infty}\frac{l_l'}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}^{\sigma}\cos m\lambda+\Delta\bar{S}_{lm}^{\sigma}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta),\end{aligned} \tag{2.43}$$

with l_l' the lateral LLN of degree l. A displacement of the Earth in eastern direction can now easily be derived by

$$\begin{aligned}\Delta e(\vartheta,\lambda) &= \frac{1}{\sin\vartheta}\frac{\partial}{\partial\lambda}\Delta\Psi(\vartheta,\lambda)\\ &= \frac{R}{\sin\vartheta}\frac{3\rho_w}{\rho_e}\sum_{l=1}^{\infty}\frac{l_l'}{2l+1}\sum_{m=0}^{l}m\cdot[-\Delta\bar{C}_{lm}^{\sigma}\sin m\lambda+\Delta\bar{S}_{lm}^{\sigma}\cos m\lambda]\bar{P}_{lm}(\cos\vartheta)\end{aligned} \tag{2.44}$$

The displacements in northern direction can be derived by

$$\begin{aligned}\Delta n(\vartheta,\lambda) &= -\frac{\partial}{\partial\vartheta}\Delta\Psi(\vartheta,\lambda)\\ &= -R\frac{3\rho_w}{\rho_e}\sum_{l=1}^{\infty}\frac{l'_l}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda+\Delta\bar{S}^{\sigma}_{lm}\sin m\lambda]\frac{\partial}{\partial\vartheta}\bar{P}_{lm}(\cos\vartheta)\end{aligned} \tag{2.45}$$

A GPS time-series $(\Delta h, \Delta e, \Delta n)$ can thus be used to estimate surface mass variations with respect to a certain reference epoch.

2.3 Gravity field determination from satellite data

The big advantage of satellite gravity data is that we can obtain a global coverage of the Earth. By tracking the orbit of a satellite, one can strengthen the estimation of the low-degree part of the gravitational field and fill in the gaps in some remote areas. Altimetry is an accurate technique to obtain an estimate of the geoid in the oceanic regions. A constellation of multiple satellites, with an inter-satellite tracking, is able to obtain the temporal changes in the gravity field and gradiometry is used to estimate the high-degree spherical harmonics of the Earth's gravitational field. In this section, we will shortly discuss the different measurement techniques used in satellite gravity modelling.

2.3.1 Satellite tracking

The orbit of a satellite is highly influenced by the gravity field of the Earth. Inversely, a time series of the position of a satellite can give extra information on this gravity field.

Satellite laser ranging (SLR)

Satellite laser ranging is the most accurate technique (within the centimetre level) to determine the position of a satellite. It measures the round trip time of pulses of light to satellites equipped with retro-reflectors. With a network of ground stations measuring the distances to the satellite, one can determine the position of the satellite's centre of mass very precisely.

Some satellites, e.g., LAGEOS-I (1976-), LAGEOS-II (1992-) and GFZ-1 (1995-1999), have been especially designed for this purpose. The shape of these passive satellites is spherical and they are fully equipped with retro-reflectors. The LAGEOS satellites fly at a high altitude $(5,900\ \mathrm{km})$, increasing their lifetime and improving the lower degrees spherical harmonics. The GFZ-1 satellite flew at a much lower altitude $(230-398\ \mathrm{km})$, reducing its lifetime considerably.

High-Low Satellite-to-Satellite Tracking (hl-SST)

An alternative method to measure the position of a satellite is to use high-low satellite-to-satellite tracking (hl-SST) by means of GPS measurements. Together with an on board accelerometer to measure the non-conservative forces (mostly atmospheric drag), this hl-SST technique is a highly accurate technique to measure the Earth's gravitational field. A recently launched satellite to have GPS receivers and a accelerometer on board is the CHAMP satellite [Reigber et al., 1999]; see also figure 2.2. The satellite was successfully launched in July 2000 with a designed lifetime of 5 years, a near polar orbit ($87°$) and an initial altitude of 454 kilometres. It was shown in Ditmar [2007] that the non-gravitational forces were only of minor influence on the gravity field determination when using the acceleration approach. In chapter 7 we will use the CHAMP hl-SST data, together with the accelerometer data, to estimate a model of the Earth's gravity field.

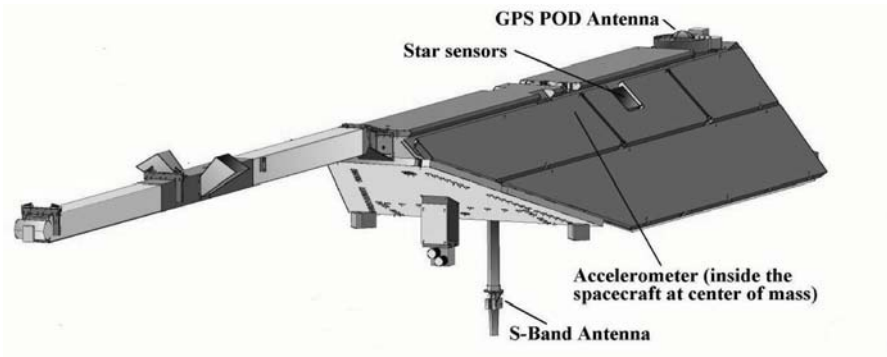


Fig. 2.2: The CHAMP satellite with the instruments for gravity field determination [GFZ]

Low-Low Satellite-to-Satellite Tracking (ll-SST)

A constellation of multiple satellites measuring their mutual distances (ll-SST), combined with an absolute positioning technique (hl-SST / SLR) can improve the estimation of the Earth's gravity field, especially its temporal variations, considerably [Jekeli, 1999]. The GRACE mission, launched in 2002, is such a mission, measuring the distance between two low-altitude ($\approx$ 500 km) satellites in identical near-polar orbits [Tapley et al., 2004a, 2004b]. The distance (on average 220 kilometres along-track) is measured by a highly accurate K-band microwave ranging system, see figure 2.3. The goal of the GRACE mission is to estimate the time-variable gravity field of the Earth with unprecedented accuracy and resolution; see, e.g., Wahr et al. [1998] and Tapley et al. [2004a].

The leader-follower configuration of GRACE is mainly sensitive along the line-of-sight, i.e., in North-South direction [Sneeuw and Schaub, 2005]. The error behaviour of

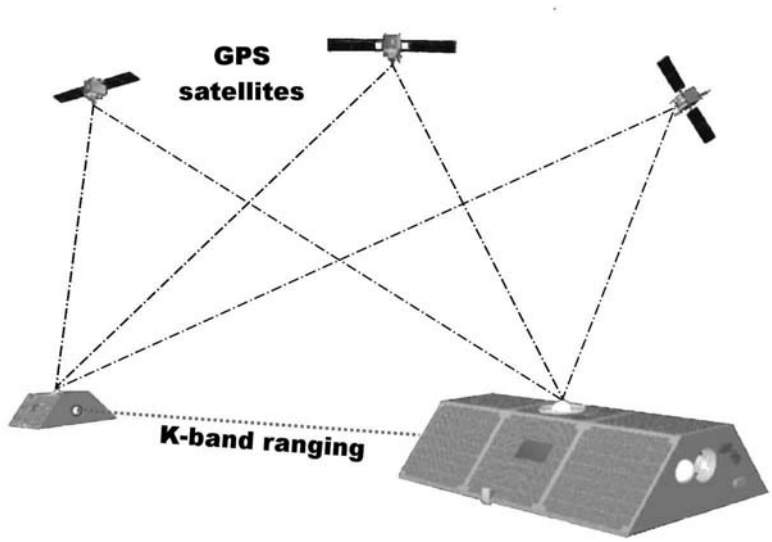


Fig. 2.3: The GRACE twin satellites [CSR]

this kind of configuration is non-isotropic. Other types of configurations (not currently in orbit) are suggested in Sneeuw et al. [2005] and Sharifi et al. [2007]. Comparisons show that the GRACE leader-follower configuration is by far not the most ideal configuration for future ll-SST missions.

2.3.2 Satellite altimetry

Satellite altimetry is a method to estimate the sea surface height, by measuring the distance between the satellite and the sea surface using an on board altimeter. Possible error sources in this height estimation are the observation error itself (including atmospheric delays), its spatial averaging and the positioning of the satellite. Moreover, one does not measure the geoid, but the sea surface height, which in general will not coincide. The difference is called *dynamic topography*. However, satellite altimetry provides very accurate data over the oceanic regions, where surface gravity data is absent. It has contributed to many combination solutions, e.g., EGM96 [Lemoine et al., 1998] and EGM08 [Pavlis et al., 2008].

Examples of altimetry satellites are the GEOS-3 (1975-1978), SEASAT (1978), GEOSAT (1985-1990), ERS-1 (1991-2000), ERS-2 (1995-), TOPEX/Poseidon (1992-2005), Jason-1 (2001-) and Envisat (2002-) satellites. Just recently (June 2008), the Jason-2 satellite is put into orbit.

2.3.3 Satellite gradiometry

Gradiometry measures the difference in acceleration of test masses over small distances with two accelerometers for each direction; see figure 2.4. In this way, components of

the gravity tensor can be measured, which will mainly provide information on the high-degree spherical harmonics of the Earth's gravity field [Rummel and Colombo, 1985]. The measured gravity gradients are expressed in units of $\mathrm{E} = 10^{-9}\mathrm{s}^{-2}$, named after the Hungarian scientist Eötvös. Non-conservative forces will act on all test masses in an almost similar way and this influence can therefore be neglected from the gradiometer observations.

The GOCE satellite will be the first satellite carrying a gradiometer on board. The launch of the satellite is scheduled for 2008, with an expected lifetime of two years. It will fly at a low-altitude (250 kilometres) and will be equipped with GPS receivers, laser reflectors (SLR), and a gradiometer. Three pairs of accelerometers will measure the gravity gradients in three directions with a baseline of approximately 0.5 metres between the corresponding accelerometers. The gradiometer noise is expected to be below $6\ \mathrm{mE}/\sqrt{\mathrm{Hz}}$ across the measurement bandwidth [Drinkwater et al., 2007]. It is expected that the GOCE-derived geoid will have an accuracy of $2\ \mathrm{cm}$ at spatial scales down to $100\ \mathrm{km}$ ($L < 200$) [Drinkwater et al., 2007].

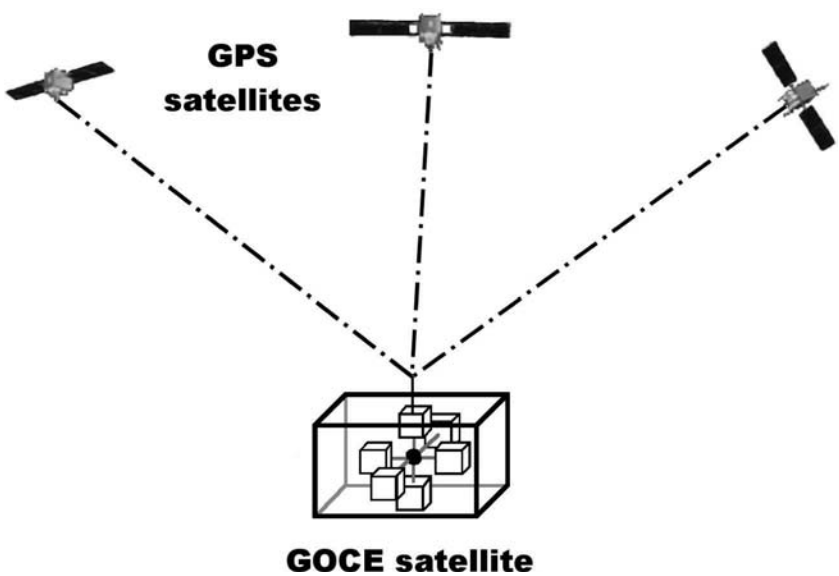


Fig. 2.4: The GOCE satellite mission

2.4 The determination of data weights in satellite gravity field modelling

In this section, we will give a historical overview of the satellite-based global gravity models. The focus of the discussion will be on the data weighting problem, i.e., how to weight the different arcs / data sets relatively to each other. This section is split into four parts: the early years (1958-1975), the period of external calibration (1975-1988),

the period of internal calibration (1988-2000), and finally the era of the dedicated space missions CHAMP, GRACE and GOCE. We have restricted ourselves to those global models where particular weight determination techniques have been applied and the most commonly used models. A more complete overview of global gravity field models can be found in Lambeck and Coleman [1983], Nerem et al. [1995], Bouman, [1998b] and the web-site of the International Center for Global Earth Models (ICGEM / GFZ).

2.4.1 The early years of satellite gravity field modelling (1958-1975)

The space age started with the launch of the first artificial satellite in 1957, the Russian *Sputnik 1*, directly followed by *Sputnik 2* and the American satellites *Explorer 1* and *Vanguard 1*. The satellites were tracked from various tracking sites to estimate the low-degree spherical harmonic coefficients. The first results on the oblateness of the Earth using artificial satellites can be found in Buchar [1958]. This marks the beginning of satellite-based gravity field modelling.

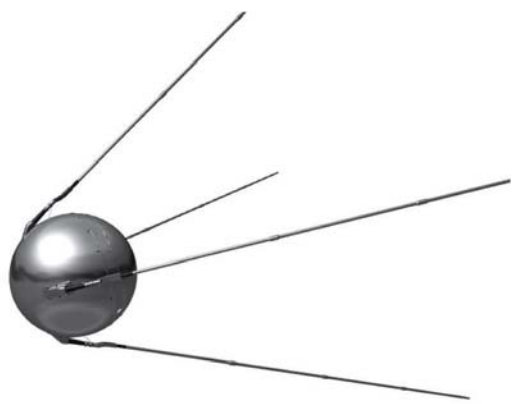


Fig. 2.5: The Sputnik 1 satellite

The first global gravity model using a combination of satellite data and surface data was computed at the Goddard Space Flight Center by W.M. Kaula [Kaula, 1961]. The satellite data represented measurements of the motion of the node, the perigee and the eccentricity of the Sputnik 2 and Vanguard 1 satellites. The surface gravity data consisted of $10^\circ \times 10^\circ$ blocks of free air anomalies as well as geoid height differences, the latter based on astro-geodetic data. The weights of the satellite data were given a-priori, confirming earlier publications about the two satellite missions; see, e.g., Merson and King-Hele [1958]. This over-weighted the satellite data [Kaula, 1961]. A least-squares adjustment was used to compute a global gravity model complete to degree 8. All variances and covariances were re-scaled a-posteriori by a single factor. This factor was computed by comparing the weighted residual square sum (neglecting the covariances) with the trace of the variance-covariance matrix of the adjusted residuals.

In the next few years, the Smithsonian Astrophysical Observatory (SOA) published some global gravity models, based on satellite tracking data and surface gravity data.

The models were named *Standard Earth (SE)*-I,-II, and -III. Several models have used the satellite tracking data and its stochastic properties from the data base of SOA, e.g., OSU-68 [Rapp, 1968] and [Koch and Morrison, 1970]. The SE-III global gravity model [Gaposchkin, 1973] used tracking data of 25 satellites. In those days, the errors in the satellite tracking data were mainly due to the bad tracking instruments on site. Therefore these measurements were grouped according to the laser instrument and were given a-priori weights, based on earlier experiences on the quality of the instruments (σ = 2-10 metres). The $5^\circ \times 5^\circ$ mean gravity anomalies were given a standard deviation based on the block size and the available data within such a block.

The potential coefficients of the satellite-only ($L = 12$) GEM-3 model were combined with $5^\circ \times 5^\circ$ mean gravity anomalies to form the ($L = 20$) combined model OSU-73 [Rapp, 1973]. The standard deviations of the GEM-3 coefficients, used in the weighting scheme of OSU-73, were estimated by comparing them to the coefficients of contemporary models. The relative weighting of the gravity anomalies with the coefficients was done by minimizing the difference between the predicted orbit of the GEOS-A satellite, based on the combined solution, and the 'true' positions, based on laser ranging data.

2.4.2 External calibration (1975-1988)

Up until 1988, with the introduction of the subset solution technique [Lerch et al., 1988], the observation weights of the satellite data were mainly based on trial and error by comparing the results to independent data like other geopotential models, satellite tracking data or altimeter data. This can be viewed as an external calibration method. Huge amounts of satellite data became available from the mid seventies. Especially the launch of the GEOS-3 satellite in 1975 and of the LAGEOS-I satellite in 1976 improved the global gravity models considerably. The altimeter on board the GEOS-3 satellite could fill in the gaps in the surface gravity data. The LAGEOS-I satellite (figure 2.6) improved the estimation of the low-degree spherical harmonics, as it was less sensible to non-gravitational forces (e.g., atmospheric drag) due to its spherical shape and its high altitude ($5,900$ km).

Fig. 2.6: The LAGEOS-I satellite

A German-French cooperation started in 1976, resulting in three global gravity models, GRIM1 [Balmino et al., 1976b], GRIM2 [Balmino et al, 1976a] and GRIM3 [Reigber et al., 1983], the latter using the new GEOS-3 altimeter data. The determination of the data weights was done rather arbitrary; e.g., the a-priori standard deviations of the surface data in GRIM3 were bounded to 5-15 mGal due to "too uneven weighting" [Reigber et al., 1983].

From 1979 to 1986, two American institutes, the Goddard Space Flight Center and the Ohio State University, based their global gravity models on the satellite-only model GEM-9 [Lerch et al., 1979], a global gravity model complete to degree 20, which was computed to improve the radial orbit determination of the GEOS-3 satellite. The GEM-9 model was combined with mean gravity anomalies to yield GEM-10 [Lerch et al., 1979], GEOS-3 altimeter data to yield GEM-10B and GEM-10C [Lerch et al., 1981], LAGEOS-I data to yeild GEM-L2 [Lerch et al., 1982,1985] and a huge amount of both surface gravity data and altimeter data to yield OSU-78 [Rapp,1978], OSU-81 [Rapp, 1981], OSU-86C/D [Rapp and Cruz, 1986a], and OSU-86E/F [Rapp and Cruz, 1986b]. The determination of the data weights in these models was done based on personal experiences and comparisons with other models. The GEOS-3 altimeter data were given a high weight to balance the influence of the continental surface gravity data (GEM-10B/C) and the LAGEOS-I data in the GEM-L2 solution had to be down-weighted by a factor of 10.

Between 1987 and 1992 a lot of effort was put into reducing the predicted radial orbit error of the future TOPEX/Poseidon altimeter satellite (launched in 1992). The requirement was to have a RMS radial error (due to gravity field mismodelling) of $10\ \mathrm{cm}$ maximum, which was clearly not met by the GEM-L2 satellite model ($65\ \mathrm{cm}$) [Nerem et al., 1993]. The Goddard Space Flight Center computed three gravity models for that purpose: GEM-T1, GEM-T2 and GEM-T3. The first model, GEM-T1, is a satellite-only solution, using both laser and optical data [Marsh et al., 1987, 1988]. The determination of the observation weights was done using trial and error by comparing the solution to independent data (external calibration). The observations were down-weighted with a factor 50, compared to a-priori zero constraints on the coefficients. Systematic errors in the laser data were probably the cause of this rather large down-weighting [Marsh et al., 1987]. The predicted radial orbit errors had a RMS of $26\ \mathrm{cm}$.

2.4.3 Internal calibration (1988-2000)

So far, any calibrations were mainly done using trial and error by comparing the results to independent data (external calibration). One of the disadvantages of this method is that the high-quality independent data were not used in the estimation of the global gravity fields. The goal was, therefore, to find a suitable *internal* calibration technique. Such a technique was proposed in Lerch et al. [1988, 1991] and compared with external calibration techniques. Different subset solutions were compared to calibrate the observation weights. For every data set, a solution was computed, in which the particular data set was omitted. The difference between the solution using all data and the solution

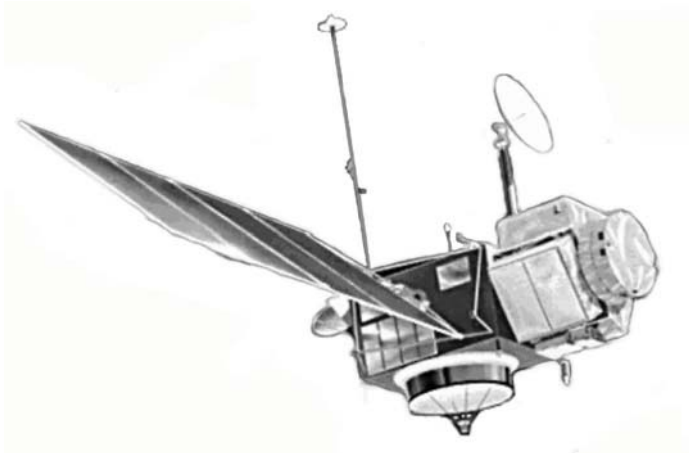


Fig. 2.7: The TOPEX-Poseidon satellite [NASA].

with the data set omitted, was compared with the stochastic model. The data set was then re-weighted and a different data set was calibrated. This algorithm was repeated until convergence. The calibrated solution was in good agreement with independent altimeter and surface gravity data. More on Lerch's subset solution method can be found in Lerch [1989, 1991] and in section 4.6.3.

The first model, that actually used the internal calibration method of Lerch was the GEM-T2 model [Marsh et al., 1989a, 1989b, 1990]. The improved data weighting and the increase in the number of observations reduced the predicted radial orbit error of TOPEX/Poseidon due to gravity from 26 to 10 cm. External calibration with independent altimeter-derived gravity anomalies showed the consistency of the new method. The GEM-T2 model was later used by the Ohio State University to compute the combined gravity field models OSU89A, OSU89B [Rapp and Pavlis, 1990], and OSU91A [Rapp et al., 1991]. The combination of the GEM-T2 model with altimeter and surface gravity data was done using external calibration.

The latest GEM-model to provide accurate radial positioning of the TOPEX/ Poseidon altimeter satellite was the GEM-T3 ($L = 50$) combined model; see, e.g., [Nerem et al., 1992] and [Lerch et al., 1994]. It has an estimated radial orbit error due to gravitational mismodelling of 6.8 cm. The satellite-only solution GEM-T3S was calibrated using the subset solution method. These relative weights were kept fixed in the combination with surface gravity data and altimeter data. SPOT 2 DORIS observations were used as an external test for the calibration method [Lerch et al., 1993]. They were later on added to the normal equations to form the GEM-T3A model [Nerem et al., 1994a]. The new model was seen as a considerable improvement, especially in the polar regions, due to the high inclination of the SPOT 2 satellite ($98.7°$). The predicted radial accuracy improved to 5 cm [Nerem et al., 1994b].

Several institutes combined their data in the mid '90s to compute new global gravity models like JGM-1 and JGM-2 [Nerem et al., 1994b], JGM-3 [Tapley et al, 1996],

GRIM-C4 [Schwintzer et al., 1997], and EGM96 [Lemoine et al., 1998]. Apart from the GRIM-C4 model, which uses external calibration techniques, the subset solution technique was used to weight the different data sets in these models. A remark could be made on the JGM-3 model, which kept the JGM-1 relative weights fixed and estimated the weights of the new data using trial and error by an external calibration to independent TOPEX/Poseidon arcs. The JGM-3 model had reduced the TOPEX/Poseidon predicted radial orbit error due to gravity field mismodelling to only $0.9\ \mathrm{cm}$ [Tapley et al., 1996].

The global gravity field models, computed at the University of Texas, like TEG-3 [Tapley et al., 1997], use a different internal calibration technique. In Yuan [1991], this method is referred to as an *optimal* weighting technique. We prefer to call it the *Iterative Maximum Likelihood Estimator (IMLE)*; see section 4.6.5. The estimator compares the residual square sum of a data set with the number of observations. The method is biased, has no minimum variance and takes no account of the degrees of freedom that are involved in estimating the vector of unknowns.

In the past few years some new Earth-Observation satellites have been launched, such as JASON-1, ENVISAT, CHAMP, and GRACE. A new global gravity model was needed as a reference field for these missions. Such a model is the GRIM5-C1 gravity model [Gruber et al, 2000]. It is a combined model of the satellite-only model GRIM5-S1 [Biancale et al., 2000] and surface data. Variance component estimation was used to compute the data weights in the satellite-only solution. Kaula constraints were added to the normal equations. Coefficients, which were determined over 90 percent by satellite data, were later used in the computation of the combined solution GRIM5-C1. The weights for the gravity anomalies were based on earlier experiences. The observation weights in the GRIM5-C1 solution were assumed as "close to optimal", as the a-posteriori variance factor approximated unity.

2.4.4 Dedicated satellite missions (2000-)

Despite several proposals in the last decades (e.g., GRAVSAT and ARISTOTELES), only recently, with the launch of the CHAMP satellite [Reigber et al., 1999] in 2000, a satellite was put into orbit with an in situ measurement system dedicated to improve the knowledge of the Earth's gravity field. In the first years after the launch of the CHAMP satellite in 2000 and the GRACE satellites in 2002, not much effort was put into the data weighting problem. Only recently, internal calibration methods, like VCE, have been used to re-weight the CHAMP [Van Loon and Kusche, 2005] and GRACE [Mayer-Gürr et al., 2007] data sets.

The first global gravity model to incorporate the new CHAMP satellite data was the EIGEN-1S model [Reigber et al., 2002], which uses three months of CHAMP satellite data, together with the normal equations of the satellite-only solution GRIM5-S1 [Biancale et al, 2000]. The relative weighting of the prior model with the new satellite data was based on external calibrations with an altimeter-derived geoid and gravity anomalies and orbit tests on different satellites [Schwintzer, personal communication]. The a-posteriori calibration of the spherical harmonic coefficients was performed by in-

dependent gravity field solutions covering different observation data periods. The key references of the CHAMP-only solutions EIGEN-2 [Reigber et al., 2003], EIGEN-3p [Reigber et al., 2005a] and TUM-1S [Gerlach et al., 2003] do not provide us with information about the estimation of the data weights. However, we can assume that the observations were equally weighted, as was also done in other CHAMP-only global gravity models, like the TUM-2Sp model [Földvary et al., 2005].

Both University of Bonn and Delft University of Technology used VCE to weight different arcs of pre-processed CHAMP data. Monte Carlo algorithms were used to speed up the computations. Our contribution can be found in Kusche and Van Loon [2005] and Van Loon and Kusche [2005] and later in this thesis, in chapter 7. Information about the CHAMP-only global gravity model of Bonn, ITG-CHAMP01, can be found in Mayer-Gürr et al. [2005] and Ilk et al. [2005].

The DEOS_CHAMP-01C_70 global gravity model [Ditmar et al., 2006] uses the second derivatives (accelerations) of the CHAMP satellite positions. The propagation of the white noise, assumed for the positions, to the acceleration observations is regularized not to overweight the low-frequency part of the noise [Ditmar and Van Eck van der Sluijs, 2004].

The first models involving GRACE observations are the American GGM01S model [Tapley et al., 2003] and the European EIGEN-GRACE01S [GFZ, 2003] and EIGEN-GRACE02S [Reigber et al., 2005b] models. We can assume that the data sets were weighted equally to estimate the global gravity models as no information is given explicitly in the corresponding references. The EIGEN-GRACE models are calibrated a-posteriori by degree using subset solutions from different time intervals. The variance by degree of the coefficients' differences between subset solutions are compared with the formal error degree variances [Reigber et al., 2005b].

The GGM02S model [Tapley et al., 2005] use the same internal calibration method (IMLE, section (4.6.5)) as was used in the TEG series; see Tapley et al. [1997] and Yuan [1991]. Three relative weights were estimated for each day, two for the GPS normal equations and one for the KBR range-rate normal equations. The K-Band range-rate had an estimated noise level of $0.4\ \mu\mathrm{m/s}$ for the earlier days of the mission and $0.2\ \mu\mathrm{m/s}$ for the data in late 2003. The GPS standard deviations ranged between 1 and $2\ \mathrm{cm}$ [Tapley et al., 2005]. The satellite-only solution was then calibrated by comparing subset solutions (internally) and by an external calibration with independent data. The calibrated model was then combined with the EGM96 solution to form the GGM02C model [Tapley et al., 2005]. However, as covariance information of EGM96 is only available up to degree 70, the VC matrix of the TEG-4 model was used instead. This model has full covariance information up to degree and order 200 and has used much the same terrestrial information as EGM96. However, down-weighting of the prior EGM96/TEG-4 model in the lower degrees and down-weighting of the GGM02S model in the higher degrees was necessary to optimally use the characteristics of both models.

A GRACE-only global gravity model, which uses Variance Components Estimation, is the ITG-GRACE02S [Mayer-Gürr et al., 2007]. Three years of GRACE data have been used to determine the satellite-only model up to degree and order 160. The data

have been divided into short arcs (30 minutes) and for each arc an individual variance component is estimated. One of the most recent GRACE models is the EIGEN-GL04C model [Förste et al, 2008]. This ($L = 360$) global gravity model uses SLR data to LAGEOS to strengthen the low-degree part of the spectrum, which was poorly solved by the GRACE satellite gravity data. Surface gravity (gravimetric and altimetric) data were used to determine the high degrees.

An even higher resolution ($L = 2160$) was obtained in the EGM08 model [Pavlis et al., 2008]. This model combines the ITG-GRACE03S GRACE-only model with a large data base of surface gravity data. It is expected that large improvements will be made up to degree and order 200 with the launch of the GOCE satellite [Drinkwater et al., 2007]. The most important satellites, weighting algorithms and global gravity models are summarized in a chronological order in table 2.2.

Tab. 2.2: Historical overview of the most important satellites, weighting algorithms and models in the field of satellite gravity field modelling

Year	Satellite	Weighting algorithm	Model
1957	Sputnik-1	Experience / trial and error	
1973			OSU-73
1975	GEOS-3	Trial and error / experience	
1976	LAGEOS-I		
1981			GEM 10B/C
1982			GEM L2
1983			GRIM-3
1988		Subset solution / trial and error	
1989			GEM-T2
1991			OSU91a
1992	TOPEX/Poseidon		
1998			EGM96
2000	CHAMP	Equal weights / VCE	GRIM5-C1
2002	GRACE		
2005			EIGEN-GRACE02S
2008			EGM08
2008	GOCE?		

2.5 Summary

This chapter gives a brief overview of physical geodesy. It explains the basics of gravity field determination, with a special focus on the use of satellite data. An overview of the determination of the data weights in satellite gravity field modelling is the topic of the second part of the chapter. One could distinguish four periods since the launch of the first satellite in 1957. The first years (1958-1975) are characterized by little

redundancy in the observations. Only a few satellites were in orbit and used for satellite gravity modelling. It was however possible to improve the estimates of the oblateness of the Earth from these first satellite observations. The data weights were chosen rather arbitrary, mostly based on the experience of the scientist or trial and error methods.

In the period 1975-1988 huge amount of gravity-related satellite observations became available, especially with the launch of the Geos-3 and the LAGEOS-I satellites. The data weights were mostly determined using external calibration techniques. The weights were chosen on a trial and error basis by comparing the results to independent data or other global gravity models.

Lerch et al. [1988] suggested to use an internal calibration technique by looking at subset solutions, omitting one observation group at a time. This method has been widely used in the period 1988-2000 to calibrate the observation groups and consequently weight the different data sets. Several combination models, such as EGM96, have used combinations of external and internal calibration to improve the stochastic model of the observations.

With the launch of the CHAMP satellite in 2000 and the GRACE satellite in 2002, a huge amount of gravity-related observations became available to be used in global gravity field modelling. Many groups computed their own global models using different functional models and different outlier treatment techniques. Most of the groups, however, did not take a proper modelling of the stochastic model into account. They considered the a-priori stochastic properties of the observations to be optimal or even used an equal data weighting approach.

3

Augmentation of the functional model

The combination of several data sets requires reliable stochastic models, the removal (or down-weighting) of outliers, and proper functional models. The functional model not only includes the (linearized) relationship between the observations and the common unknown parameters, but may also include some local parameters, which model systematic effects or inconsistencies among the data sets. This chapter starts with an introduction to the least-squares approach. Hypothesis testing on the significance of certain local parameters is the subject of the second part. The chapter ends with the Best Linear Softly Unbiased Estimator (BLSUE), which weakens the constraints of unbiasedness in the combination of new data with prior information.

3.1 The Least-Squares principle

3.1.1 The Least-Squares estimator of the vector of unknowns

We assume a general linear Gauss-Markov Model as the functional model, i.e.,

$$\underline{y} = A\underline{x} + \underline{e} \tag{3.1}$$

where

$\underline{y}$ = $m \times 1$ vector of observations
A = $m \times n$ design matrix
$\underline{x}$ = $n \times 1$ vector of unknowns
$\underline{e}$ = $m \times 1$ vector of residuals.

We have chosen to use this Gauss-Markov model throughout the thesis. Other functional models, like the mixed linear model or the conditional model, can always be re-written as a Gauss-Markov model.

The stochastic properties of the residuals are described by the stochastic model:

$$E\{\underline{e}\} = \underline{0} \quad ; \quad D\{\underline{e}\} = Q_y \tag{3.2}$$

where $E\{.\}$ denotes the expectation operator, and $D\{.\}$ the dispersion operator, defined as

$$D\{\underline{e}\} := E\{(\underline{e} - E\{\underline{e}\})(\underline{e} - E\{\underline{e}\})^T\}. \tag{3.3}$$

The observations can take any value in the $\mathbb{R}^m$ space due to observation errors. The observations are, however, *expected* to lie in $\mathcal{R}(A)$, the column space of A, since $E\{\underline{y}\} = A\underline{x}$. Any misfit is expressed in the vector $\underline{e}$. As a result, the variance-covariance matrix of the observations Q_y equals the variance-covariance matrix of the residuals. In a *weighted least-squares* approach (see e.g., Teunissen [2000a]), the weighted residual square sum Ω is minimized, where Ω is defined as

$$\Omega := \| \underline{e} \|^2_{Q_y} = \underline{e}^T Q_y^{-1} \underline{e}. \tag{3.4}$$

Minimization of Ω by varying $\underline{x}$ results in the least-squares estimate $\underline{\hat{x}}$ of the vector of unknowns $\underline{x}$:

$$\hat{\Omega} := \min_{\underline{x}} \underline{e}^T Q_y^{-1} \underline{e} = \min_{\underline{x}} (\underline{y} - A\underline{x})^T Q_y^{-1} (\underline{y} - A\underline{x}) = (\underline{y} - A\underline{\hat{x}})^T Q_y^{-1} (\underline{y} - A\underline{\hat{x}}). \tag{3.5}$$

The vector $\underline{\hat{x}}$ is the solution of the system of normal equations

$$N\underline{\hat{x}} = \underline{h}, \tag{3.6}$$

with N the normal matrix and $\underline{h}$ the right-hand-side vector defined as [Teunissen, 2000a]

$$\begin{aligned} N &:= A^T Q_y^{-1} A \\ \underline{h} &:= A^T Q_y^{-1} \underline{y}, \end{aligned}$$

The least-squares estimate $\underline{\hat{x}}$ has maximum likelihood and minimum variance if one assumes a normal distribution with known functional and stochastic model. The propagation law of variances is used to derive the estimator of the variance-covariance matrix of $\underline{\hat{x}}$

$$Q_{\hat{x}} = N^{-1} A^T Q_y^{-1} Q_y Q_y^{-1} A N^{-1} = N^{-1} N N^{-1} = N^{-1}. \tag{3.7}$$

The vector $\underline{\hat{y}} := A\underline{\hat{x}}$ is the least-squares estimator of the vector of observations $\underline{y}$ and can be computed by:

$$\underline{\hat{y}} = A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} \underline{y} = P_A \underline{y}. \tag{3.8}$$

where P_A is a projection matrix, which projects a vector onto the column space $\mathcal{R}(A)$ and along its orthogonal complement (in Q_y-space). The least-squares estimator of the residual vector is the difference between $\underline{y}$ and $\underline{\hat{y}}$:

$$\underline{\hat{e}} = [I - A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1}] \underline{y} = P_A^{\perp} \underline{y}, \tag{3.9}$$

with $P_A^{\perp}$ the projection matrix, which projects *along* the column space $\mathcal{R}(A)$ and *onto* its orthogonal complement. Useful properties of the aforementioned projection matrices are [Strang, 1988]:

1. P_A, $P_A^\perp$ are symmetric $m \times m$ matrices if Q_y is a scaled identity matrix.

2. P_A, $P_A^\perp$ are idempotent matrices, i.e. $P_A P_A = P_A$ and $P_A^\perp P_A^\perp = P_A^\perp$.

3. $P_A = I - P_A^\perp$ and $P_A^\perp = I - P_A$.

4. $P_A^T Q_y^{-1} = P_A^T Q_y^{-1} P_A = Q_y^{-1} P_A$ and $(P_A^\perp)^T Q_y^{-1} = (P_A^\perp)^T Q_y^{-1} P_A^\perp = Q_y^{-1} P_A^\perp$.

5. $\mathcal{R}(P_A) = \mathcal{R}(A) = \mathcal{N}(P_A^\perp)$ and $\mathcal{R}(P_A^\perp) = \mathcal{R}(A)^\perp = \mathcal{N}(P_A)$ with $\mathcal{R}(\cdot)$ the range operator and $\mathcal{N}(\cdot)$ the null space operator.

3.1.2 The overall model test

We have now derived an estimator for the vector of unknowns using the least-squares approach (Eq. (3.6)) and we have an estimate of the variance-covariance matrix of this estimator (Eq. (3.7)). If the design matrix A is of full column rank and the number of observations is at least equal to the number of unknowns, one will always find a solution of the vector of unknowns $\underline{x}$. However, the residuals can be relatively large, if one would compare them to the stochastic properties of the observations, defined by the stochastic model. One should therefore *test* if the observations fit the functional and stochastic model. Two hypotheses can be tested against each other:

$$H_0 : E\{\underline{y}\} = A\underline{x} \quad ; \quad H_A : E\{\underline{y}\} = A\underline{x} + B\underline{\delta}, \tag{3.10}$$

with $B\underline{\delta}$ a matrix-vector product, which can account for any misfit in the data. The test can be used to test for any bias in the data, i.e. $B = [\ 1 \ \ 1 \ \ \dots \ \ 1\ ]^T$ or to test for a specific outlier $B = [\ 0 \ \ \dots \ \ 0 \ \ 1 \ \ 0 \ \ \dots \ \ 0\ ]^T$. In both cases, the matrix B becomes a vector and $\underline{\delta}$ a scalar. If one increases the dimensions of B to its maximum $m \times (m-n)$, the expectation $E\{\underline{y}\}$ can take any value within $\mathbb{R}^m$. A suitable test statistic for this problem is the quadratic form $\hat{\Omega}$ [Teunissen, 2000b], see Eq. (3.5), which is distributed under H_0 and H_A as

$$H_0 : \hat{\Omega} \sim \chi^2(m-n, 0) \quad ; \quad H_A : \hat{\Omega} \sim \chi^2(m-n, \lambda) \tag{3.11}$$

with the non-centrality parameter defined as

$$\lambda := \underline{\delta}^T B^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} B\underline{\delta}. \tag{3.12}$$

Note that the matrix Q_y is assumed to be known, which is in general not valid for real data. Equivalent to this test statistic, one could also consider the unbiased estimator of the variance factor of unit weight (or a posterior variance factor):

$$\hat{\sigma}^2 = \frac{\hat{\Omega}}{m-n}, \tag{3.13}$$

which is distributed under H_0 and H_A as

$$H_0 : \hat{\sigma}^2 \sim F(m-n, \infty, 0) \quad ; \quad H_A : \hat{\sigma}^2 \sim F(m-n, \infty, \lambda) \tag{3.14}$$

This test is called *the overall model test* or *the global model test*. It compares $\hat{\Omega}$ with its expectation under H_0, i.e.

$$E\{\hat{\Omega}\} = m - n. \tag{3.15}$$

Rejection of the H_0-hypothesis implies either

- outlier(s) in the data (error in $\underline{y}$),
- unmodelled systematic effects in (parts of) the data (error in $A\underline{x}$),
- or a wrong variance-covariance matrix (error in Q_y).

In this chapter, we will focus on the augmentation of the functional model, i.e., on redefining the design matrix A. The augmentation of the stochastic model and the detection of outliers will be discussed in chapters 4 and 6.

3.2 Global and local parameters

3.2.1 The estimation of global and local parameters

It is common practice in gravity field modelling to use different satellite data sets in a common least-squares adjustment, e.g., data from different missions, satellite gravity data from different periods (e.g., individual arcs) within one mission, a combination with terrestrial or altimetric data or a combination with an a-priori model. Suppose, the noise of these data sets can be considered to be mutually uncorrelated. The stochastic model can then be written as

$$E\{\begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_p \end{bmatrix}\} = \underline{0} \quad ; \quad D\{\begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_p \end{bmatrix}\} = \begin{bmatrix} W_1^{-1} & 0 & \cdots & 0 \\ 0 & W_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W_p^{-1} \end{bmatrix} \tag{3.16}$$

The general case of a functional model, in which we can assume different observation groups, reads

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_p \end{bmatrix} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix} \underline{x} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_p \end{bmatrix}. \tag{3.17}$$

We will call such a model *a functional model of type I*. The least-squares estimate follows from solving Eq. (3.6), in which the normal matrix N and the right-hand-side vector $\underline{h}$ are defined by

$$\begin{aligned} N &= \sum_{i=1}^{p} N_i \quad ; \quad N_i = A_i^T W_i A_i; \\ \underline{h} &= \sum_{i=1}^{p} \underline{h}_i \quad ; \quad \underline{h}_i = A_i^T W_i \underline{y}_i. \end{aligned} \tag{3.18}$$

In general, the observations are not only dependent on the common or *global* parameters, such as spherical harmonic coefficients. Inconsistencies can occur in the functional model, such as datum inconsistencies [Fotopoulos, 2005], systematic biases [Kotsakis, 2005] and geographically correlated errors. These effects can be modelled by *local* parameters. Dividing the vector of unknowns in global and local parameters can reduce the computational effort of the least-squares solution considerably. The *functional model of type II* can be written as

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_p \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 & A_{1s} \\ 0 & A_{22} & \cdots & 0 & A_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{pp} & A_{ps} \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_p \\ \underline{x}_s \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_p \end{bmatrix} \tag{3.19}$$

The normal equations will reduce to (see, e.g., Lucas and Dillinger [1998])

$$\begin{bmatrix} N_{11} & 0 & \cdots & 0 & N_{1s} \\ 0 & N_{22} & \cdots & 0 & N_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & N_{pp} & N_{ps} \\ N_{s1} & N_{s2} & \cdots & N_{sp} & N_{ss} \end{bmatrix} \begin{bmatrix} \hat{\underline{x}}_1 \\ \hat{\underline{x}}_2 \\ \vdots \\ \hat{\underline{x}}_p \\ \hat{\underline{x}}_s \end{bmatrix} = \begin{bmatrix} \underline{h}_{11} \\ \underline{h}_{22} \\ \vdots \\ \underline{h}_{pp} \\ \underline{h}_s \end{bmatrix}. \tag{3.20}$$

The sub-matrices ($i = 1, \ldots, p$) can be computed by

$$\begin{aligned} N_{ii} &= A_{ii}^T W_i A_{ii} \\ N_{is} &= A_{ii}^T W_i A_{is} \\ N_{si} &= N_{is}^T \\ N_{ss} &= A_{1s}^T W_1 A_{1s} + A_{2s}^T W_2 A_{2s} + \ldots + A_{ps}^T W_p A_{ps} = \sum_{i=1}^{p} A_{is}^T W_i A_{is} \end{aligned} \tag{3.21}$$

and the right-hand-side vectors by

$$\begin{aligned} \underline{h}_{ii} &= A_{ii}^T W_i \underline{y}_i \\ \underline{h}_s &= A_{1s}^T W_1 \underline{y}_1 + A_{2s}^T W_2 \underline{y}_2 + \ldots + A_{ps}^T W_p \underline{y}_p = \sum_{i=1}^{p} A_{is}^T W_i \underline{y}_i. \end{aligned} \tag{3.22}$$

If the functional model is written as a functional model of type III, the system of normal equations to solve for the global parameters $\underline{x}_s$ can be reduced to (see, e.g., Reigber [1989]):

$$\bar{N}_{ss}\hat{\underline{x}}_s = \bar{\underline{h}}_s, \tag{3.23}$$

where

$$\begin{aligned} \bar{N}_{ss} &= N_{ss} - \sum_{i=1}^{p} N_{si}N_{ii}^{-1}N_{is} \\ \bar{\underline{h}}_s &= \underline{h}_s - \sum_{i=1}^{p} N_{si}N_{ii}^{-1}\underline{h}_{ii}, \end{aligned} \tag{3.24}$$

or [Lucas and Dillinger, 1998]

$$\begin{aligned} \bar{N}_{ss} &= \sum_{i=1}^{p} \Upsilon_i \quad ; \quad \Upsilon_i := A_{is}^T W_i A_{is} - N_{is}^T N_{ii}^{-1} N_{is} \\ \bar{\underline{h}}_s &= \sum_{i=1}^{p} \underline{\psi}_i \quad ; \quad \underline{\psi}_i := A_{is}^T W_i \underline{y}_i - N_{is}^T N_{ii}^{-1} \underline{h}_{ii}. \end{aligned} \tag{3.25}$$

The variance-covariance matrix of the global parameters then reads

$$Q_{\hat{\underline{x}}_s} = \bar{N}_{ss}^{-1}. \tag{3.26}$$

The least-squares solution of the local parameters $\underline{x}_i$ can be computed by using the *back-substitution step*

$$\begin{aligned} N_{ii}\hat{\underline{x}}_i &= \underline{h}_{ii} - N_{is}\hat{\underline{x}}_s \\ &= A_{ii}^T W_i(\underline{y}_i - A_{is}\hat{\underline{x}}_s) \end{aligned} \tag{3.27}$$

with variance-covariance matrix

$$Q_{\hat{\underline{x}}_i} = N_{ii}^{-1} + N_{ii}^{-1}N_{is}\bar{N}_{ss}^{-1}N_{si}N_{ii}^{-1}. \tag{3.28}$$

Instead of solving a large system of normal equations with a $n \times n$ normal matrix, now several smaller systems need to be solved with normal matrices of varying size. As the computational cost increases dramatically with the dimension of the normal matrix (e.g., Cholesky factorization: n^3), the division of the vector of unknowns into global and local parameters will reduce the computation time considerably.

3.2.2 Test of significance for local parameters

Naturally, by adding an increasing number of local parameters, we can improve the fit of the data to the model. However, one should first test any augmentation of the functional model. Consider a functional model of type II, in which the local parameters of group

k are split into two parts, i.e. $\underline{x}_{k,1}$ and $\underline{x}_{k,2}$:

$$\begin{bmatrix} \underline{y}_1 \\ \underline{y}_2 \\ \vdots \\ \underline{y}_k \\ \vdots \\ \underline{y}_p \end{bmatrix} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 & A_{1s} \\ 0 & A_{22} & \cdots & 0 & 0 & \cdots & 0 & A_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & A_{kk,1} & A_{kk,2} & \cdots & 0 & A_{ks} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & A_{pp} & A_{ps} \end{bmatrix} \begin{bmatrix} \underline{x}_1 \\ \underline{x}_2 \\ \vdots \\ \underline{x}_{k,1} \\ \underline{x}_{k,2} \\ \vdots \\ \underline{x}_p \\ \underline{x}_s \end{bmatrix} + \begin{bmatrix} \underline{e}_1 \\ \underline{e}_2 \\ \vdots \\ \underline{e}_k \\ \vdots \\ \underline{e}_p \end{bmatrix} \tag{3.29}$$

We will now test the significance of $\underline{x}_{k,2}$. Two hypotheses are tested against each other:

$$H_0 : \underline{x}_{k,2} = \underline{0} \quad ; \quad H_A : \underline{x}_{k,2} \neq \underline{0}, \tag{3.30}$$

When accepting the H_0-hypothesis, the vector $\underline{x}_{k,2}$ is not significant and should be removed from the vector of unknowns. The system of normal equations then reads

$$\begin{bmatrix} N_{11} & \cdots & 0 & 0 & 0 & \cdots & 0 & N_{1s} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & N_{k-1,k-1} & 0 & 0 & \cdots & 0 & N_{k-1,s} \\ 0 & \cdots & 0 & N_{kk,11} & 0 & \cdots & 0 & N_{ks,1} \\ 0 & \cdots & 0 & 0 & N_{k+1,k+1} & \cdots & 0 & N_{k+1,s} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & N_{pp} & N_{ps} \\ N_{s1} & \cdots & N_{s,k-1} & N_{sk,1} & N_{s,k+1} & \cdots & N_{sp} & N_{ss} \end{bmatrix} \begin{bmatrix} \hat{\underline{x}}_1 \\ \vdots \\ \hat{\underline{x}}_{k-1} \\ \hat{\underline{x}}_{k,1} \\ \hat{\underline{x}}_{k+1} \\ \vdots \\ \hat{\underline{x}}_p \\ \hat{\underline{x}}_s \end{bmatrix}_0 = \begin{bmatrix} \underline{h}_{11} \\ \vdots \\ \underline{h}_{k-1k-1} \\ \underline{h}_{kk,1} \\ \underline{h}_{k+1k+1} \\ \vdots \\ \underline{h}_{pp} \\ \underline{h}_s \end{bmatrix}. \tag{3.31}$$

with

$$\begin{aligned} N_{kk,11} &= A_{kk,1}^T W_k A_{kk,1} \\ N_{ks,1} &= A_{kk,1}^T W_k A_{ks} \\ N_{sk,1} &= N_{ks,1}^T \\ \underline{h}_{kk,1} &= A_{kk,1}^T W_k \underline{y}_k. \end{aligned} \tag{3.32}$$

This can be re-written as

$$N_0 \hat{\underline{x}}_0 = \underline{h}_0. \tag{3.33}$$

Note that the removal of $\underline{x}_{k,2}$ from the vector of unknowns does not change N_{ss} and $\underline{h}_s$. Rejecting the H_0-hypothesis implies accepting the significance of the local parameters

$\underline{x}_{k,2}$. The normal equations extend to

$$\left[\begin{array}{c|cccccc} N_{kk,22} & 0 & \cdots & N_{kk,21} & \cdots & 0 & N_{ks,2} \\ \hline 0 & N_{11} & \cdots & 0 & \cdots & 0 & N_{1s} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ N_{kk,12} & 0 & \cdots & N_{kk,11} & \cdots & 0 & N_{ks,1} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & N_{pp} & N_{ps} \\ N_{sk,2} & N_{s1} & \cdots & N_{sk,1} & \cdots & N_{sp} & N_{ss} \end{array}\right] \left[\begin{array}{c} \underline{\hat{x}}_{k,2} \\ \hline \underline{\hat{x}}_1 \\ \vdots \\ \underline{\hat{x}}_{k,1} \\ \vdots \\ \underline{\hat{x}}_p \\ \underline{\hat{x}}_s \end{array}\right]_0 = \left[\begin{array}{c} \underline{h}_{kk,2} \\ \hline \underline{h}_{11} \\ \vdots \\ \underline{h}_{kk,1} \\ \vdots \\ \underline{h}_{pp} \\ \underline{h}_s \end{array}\right]. \tag{3.34}$$

with

$$\begin{array}{ll} N_{kk,ij} & = A_{kk,i}^T W_k A_{kk,j} \\ N_{ks,2} & = A_{kk,2}^T W_k A_{ks} \\ N_{sk,2} & = N_{ks,2}^T \\ \underline{h}_{kk,2} & = A_{kk,2}^T W_k \underline{y}_k. \end{array} \tag{3.35}$$

The system of normal equations can be re-written as

$$\left[\begin{array}{cc} N_{kk,22} & N_{k2,0} \\ N_{0,k2} & N_0 \end{array}\right] \left[\begin{array}{c} \underline{\hat{x}}_{k,2} \\ \underline{\hat{x}}_A \end{array}\right] = \left[\begin{array}{c} \underline{h}_{kk,2} \\ \underline{h}_0 \end{array}\right]. \tag{3.36}$$

The local parameters $\underline{\hat{x}}_{k,2}$ can now be retrieved after reduction of the normal equations:

$$\left[N_{kk,22} - N_{k2,0} N_0^{-1} N_{k2,0}\right] \underline{\hat{x}}_{k,2} = \underline{h}_{kk,2} - N_{k2,0} \underline{\hat{x}}_0. \tag{3.37}$$

Using the back-substitution step, we find

$$\underline{\hat{x}}_0 - \underline{\hat{x}}_A = N_0^{-1} N_{0,k2} \underline{\hat{x}}_{k,2}. \tag{3.38}$$

The two hypotheses can be tested by analyzing the difference

$$\delta\hat{\Omega} := \hat{\Omega}_0 - \hat{\Omega}_A \tag{3.39}$$

which is distributed as

$$H_0 : \delta\hat{\Omega} \sim \chi^2(n_{k,2}) \quad ; H_A : \delta\hat{\Omega} \sim \chi^2(n_{k,2}, \lambda), \tag{3.40}$$

where λ is defined by Eq. (3.42). The test statistic $\delta\hat{\Omega}$ can be computed as

$$\begin{array}{ll} \delta\hat{\Omega} & = (\underline{\hat{x}}_A - \underline{\hat{x}}_0)^T \underline{h}_0 + \underline{\hat{x}}_{k,2}^T \underline{h}_{kk,2} \\ & = -\underline{\hat{x}}_{k,2}^T N_{k2,0} N_0^{-1} \underline{h}_0 + \underline{\hat{x}}_{k,2}^T \underline{h}_{kk,2} \\ & = \underline{\hat{x}}_{k,2}^T \left[N_{kk,22} - N_{k2,0} N_0^{-1} N_{k2,0}\right] \underline{\hat{x}}_{k,2} \\ & = \underline{\hat{x}}_{k,2}^T Q_{\underline{\hat{x}}_{k,2}}^{-1} \underline{\hat{x}}_{k,2} \end{array} \tag{3.41}$$

This is in agreement with Schaffrin and Iz [2001]. As $E\{\underline{\hat{x}}_{k,2}\} = \underline{x}_{k,2}$ and $D\{\underline{\hat{x}}_{k,2}\} = [N_{kk,22} - N_{k2,0} N_0^{-1} N_{k2,0}]^{-1}$, the non-centrality parameter reads

$$\lambda = \underline{x}_{k,2}^T \left[N_{kk,22} - N_{k2,0} N_0^{-1} N_{k2,0}\right] \underline{x}_{k,2}^T. \tag{3.42}$$

Rejecting H_0 implies accepting H_A, which justifies the addition of the tested (local) parameters $\underline{x}_{k,2}$ to the vector of unknowns.

3.3 The addition of prior information

3.3.1 Augmentation of the functional model

The addition of prior information on any linear combination of the unknown parameters leads to an augmentation of the functional model:

$$\begin{bmatrix} \underline{y} \\ \underline{y}_P \end{bmatrix} = \begin{bmatrix} A \\ H \end{bmatrix} \underline{x}_0 + \begin{bmatrix} \underline{e} \\ \underline{e}_P \end{bmatrix}, \tag{3.43}$$

where $\underline{y}_P$ is the vector of prior information on the vector of unknowns $\hat{\underline{x}}$, i.e., $\underline{y}_P = H\hat{\underline{x}}_P$, where H is a $q \times n$ design matrix of the prior information, containing the q linear combinations (assuming rank H=q). The corresponding stochastic model reads

$$E\{\begin{bmatrix} \underline{e} \\ \underline{e}_P \end{bmatrix}\} = \underline{0} \quad ; \quad D\{\begin{bmatrix} \underline{e} \\ \underline{e}_P \end{bmatrix}\} = \begin{bmatrix} Q_y & 0 \\ 0 & HN_P^{-1}H^T \end{bmatrix}, \tag{3.44}$$

with N_P the inverse of the variance-covariance matrix of the prior information $\hat{\underline{x}}_P$. Such a linear combination $H\hat{\underline{x}}_P$ could represent a variety of information, e.g., geoid information in a certain area or information about certain spherical harmonic degrees. The least-squares solution reads

$$\begin{aligned} \hat{\underline{x}}_0 &= (N + H^T(HN_P^{-1}H^T)^{-1}H)^{-1}(A^TQ_y^{-1}\underline{y} + H^T(HN_P^{-1}H^T)^{-1}\underline{y}_P) \\ &= \hat{\underline{x}}_A + (N + H^T(HN_P^{-1}H^T)^{-1}H)^{-1}H^T(HN_P^{-1}H^T)^{-1}(\underline{y}_P - H\hat{\underline{x}}_A), \end{aligned} \tag{3.45}$$

where $\hat{\underline{x}}_A$ is the least-squares estimate if only the satellite data is used, i.e.,

$$\underline{y} = A\underline{x}_A + \underline{e}, \tag{3.46}$$

with $E\{\underline{e}\} = \underline{0}$ and $D\{\underline{e}\} = Q_y$.

3.3.2 Test for consistency with prior information

The prior information $\underline{y}_P$ should be tested for consistency with the satellite data $\underline{y}$ before combining the two data sets in a common least-squares adjustment. Two hypotheses will therefore be tested against each other:

$$H_0 : E\{\underline{y}_P\} = E\{H\hat{\underline{x}}_A\} \quad ; \quad H_A : E\{\underline{y}_P\} \neq E\{H\hat{\underline{x}}_A\}. \tag{3.47}$$

If the H_0-hypothesis is valid, the two data sets $\underline{y}$ and $\underline{y}_P$ are consistent with each other and one should use the functional model of Eq. (3.43). If this hypothesis is not valid, one should only use the 'old' data set $\underline{y}$, see Eq. (3.46) or augment the functional model. The two hypotheses can be tested by evaluating the increase in the residual square sum when assuming consistency between the two data sets. Therefore, the difference between $\hat{\Omega}_0$ and $\hat{\Omega}_A$ needs to be computed, where

$$\hat{\Omega}_0 = (\underline{y} - A\hat{\underline{x}}_0)^TQ_y^{-1}(\underline{y} - A\hat{\underline{x}}_0) + (\hat{\underline{x}}_P - \hat{\underline{x}}_0)^TH^T(HN_P^{-1}H^T)^{-1}H(\hat{\underline{x}}_P - \hat{\underline{x}}_0), \tag{3.48}$$

and

$$\hat{\Omega}_A = (\underline{y} - A\underline{\hat{x}}_A)^T Q_y^{-1} (\underline{y} - A\underline{\hat{x}}_A). \tag{3.49}$$

The increase then reads (cf. [Schaffrin, 1987])

$$\begin{aligned} \delta\hat{\Omega} &:= \hat{\Omega}_0 - \hat{\Omega}_A \\ &= -(\underline{\hat{x}}_0 - \underline{\hat{x}}_A)^T A^T Q_y^{-1} \underline{y} + \underline{\hat{x}}_P^T H^T (H N_P^{-1} H^T)^{-1} H (\underline{\hat{x}}_P - \underline{\hat{x}}_0) \\ &= (\underline{\hat{x}}_P - \underline{\hat{x}}_A)^T H^T (H N^{-1} H^T + H N_P^{-1} H^T)^{-1} H (\underline{\hat{x}}_P - \underline{\hat{x}}_A), \end{aligned} \tag{3.50}$$

which is distributed as

$$H_0' : \delta\hat{\Omega} \sim \chi^2(q) \quad ; \quad H_A' : \delta\hat{\Omega} \sim \chi^2(q, \lambda), \tag{3.51}$$

with

$$\lambda = \underline{\delta}^T H^T (H N^{-1} H^T + H N_P^{-1} H^T)^{-1} H \underline{\delta}. \tag{3.52}$$

The vector $\underline{\delta}$ is defined as the discrepancy between the expectation of $\underline{\hat{x}}_P$ and $\underline{\hat{x}}_A$, i.e.

$$\underline{\delta} := E\{\underline{\hat{x}}_P\} - E\{\underline{\hat{x}}_A\}, \tag{3.53}$$

which is assumed to be non-zero under H_A. Finally, the test statistic for the addition of prior information to the functional model reads:

$$\hat{T} := \frac{\delta\hat{\Omega}}{q} / \frac{\hat{\Omega}_A}{m-n} \tag{3.54}$$

which is distributed as

$$H_0'' : \hat{T} \sim F(q, m-n, 0) \quad ; \quad H_A'' : \hat{T} \sim F(q, m-n, \lambda). \tag{3.55}$$

The acceptatance of H_0, H_0' or H_0'' implies consistency between the new data and the prior information. When the hypothesis is rejected, there may exist an inconsistency between both data sets, which can often be modelled by adding extra parameters to the functional model.

3.4 Softly unbiased estimation

One of the disadvantages of the least-squares combination is that the solution $\underline{\hat{x}}_0$ is sensitive to errors in the prior information. Schaffrin [1985, 1986, 1987] tried to minimize this sensitivity by introducing an alternative combination method, which he called *robust collocation*. In Middel and Schaffrin [1987] this method was applied to global gravity field modelling. The robust collocation was first developed for the mixed linear model. The equivalent estimator for the Gauss-Markov model is called the *Best Linear Softly Unbiased Estimator* (BLSUE) [Schaffrin, 1999]. The BLSUE is a mixture between the Best Linear Unbiased Estimator (BLUE) and the Best Linear Estimator. First, we will discuss the BLUE and the BLE and then derive the equations for the BLSUE.

Best Linear Unbiased Estimator (BLUE)

We start with the Gauss-Markov model:

$$\underline{y} = A\underline{x} + \underline{e} \quad ; \quad E\{\underline{e}\} = \underline{0} \quad ; \quad D\{\underline{e}\} = Q_y. \tag{3.56}$$

The Best Linear Unbiased Estimator (BLUE) minimizes the trace of the Mean Square Error (MSE) matrix under the condition of unbiasedness. This matrix reads

$$MSE\{\underline{\hat{x}}\} = D\{\underline{\hat{x}}\} + (E\{\underline{\hat{x}}\} - \underline{x})(E\{\underline{\hat{x}}\} - \underline{x})^T \tag{3.57}$$

and is equal to $D\{\underline{\hat{x}}\}$ if we can assume unbiasedness, i.e., $E\{\underline{\hat{x}}\} = \underline{x}$. Writing $\underline{\hat{x}} = F\underline{y}$, the BLUE can be achieved by minimizing the trace of $D\{\underline{\hat{x}}\} = FQ_y^{-1}F^T$ under the unbiasedness constraint $FA = I_n$. Note that $E\{\underline{\hat{x}}\} = FE\{\underline{y}\} = FA\underline{x} = \underline{x} \rightarrow FA = I_n$. The BLUE can be found by minimizing the Lagrange function

$$\Psi_{BLUE}(F, \Lambda) = \text{tr}(FQ_yF^T) + 2\text{tr}(\Lambda(FA - I)) \tag{3.58}$$

The minimum is attained if

$$\underline{\hat{x}}_{BLUE} = (A^TQ_y^{-1}A)^{-1}A^TQ_y^{-1}\underline{y} = N^{-1}\underline{h}, \tag{3.59}$$

which is the well-known weighted least-squares equation.

Best Linear Estimator (BLE)

The Best Linear Estimator (BLE) minimizes $tr(MSE\{\underline{\hat{x}}\})$, but does not use the assumption of unbiasedness. Therefore,

$$\begin{aligned} MSE\{\underline{\hat{x}}\} &= D\{\underline{\hat{x}}\} + (E\{\underline{\hat{x}}\} - \underline{x})(E\{\underline{\hat{x}}\} - \underline{x})^T \\ &= D\{\underline{\hat{x}}\} + (FA\underline{x} - \underline{x})(FA\underline{x} - \underline{x})^T. \end{aligned} \tag{3.60}$$

The BLE objective function reads

$$\Psi_{BLE}(F) = \text{tr}(FQ_yF^T) + \text{tr}((FA - I)\underline{x}\underline{x}^T(FA - I)^T). \tag{3.61}$$

Minimizing this function leads to the equation of the BLE:

$$\underline{\hat{x}}_{BLE} = (I_n + \underline{x}\underline{x}^TN)^{-1}\underline{x}\underline{x}^T\underline{h}, \tag{3.62}$$

or, using Eq. (B.6),

$$\underline{\hat{x}}_{BLE} = \underline{x} \cdot \frac{\underline{x}^T\underline{h}}{1 + \underline{x}^TN\underline{x}}; \tag{3.63}$$

see also Schaffrin [2007]. As the vector $\underline{x}$ is not known, one could replace this vector by the estimate $\underline{x}_{BLE}$ and iterate till convergence. This estimate is called *repro*-BLE [Schaffrin, 2007]. It was also suggested in the same paper to replace $\underline{x}\underline{x}^T$ by $\alpha^{-1} \cdot S$, resulting in the equation for the so-called α-weighted S-homBLE:

$$\underline{\hat{x}}_{homBLE} = (N + \alpha S^{-1})^{-1}\underline{h}, \tag{3.64}$$

which can be seen as a Tikhonov-Philips regularization.

Best Linear Softly Unbiased Estimator (BLSUE)

The Best Linear Softly Unbiased Estimator (BLSUE) adds to the target function of BLE the constraint of softly unbiasedness:

$$FA = I_n + E_0^T \quad ; \quad vec(E_0) \sim (\underline{0}, I_n \otimes Q_0) \tag{3.65}$$

Hence, the uncertainty of the unbiasedness is expressed in the matrix Q_0. The target function of BLSUE then reads [Schaffrin, 1995]

$$\Psi_{BLSUE}(F,\Lambda) = \text{tr}(FQ_yF^T)+\text{tr}((FA-I)\underline{x}\underline{x}^T(FA-I)^T)+2\text{tr}(\Lambda(FA-I))-\text{tr}(\Lambda^TQ_0\Lambda) \tag{3.66}$$

For $Q_0 \to 0$ the BLSUE converges to the BLUE, for $Q_0 \to \infty$ it will be equal to the BLE. Minimizing this target function yields [Schaffrin, 1999]

$$\underline{\hat{x}}_{BLSUE} = (I_n + (\underline{x}\underline{x}^T + Q_0^{-1})N)^{-1}(\underline{x}\underline{x}^T + Q_0^{-1})\underline{h} \tag{3.67}$$

Using Eq. (B.6) this can be re-written as

$$\underline{\hat{x}}_{BLSUE} = [N + Q_0 - (1 + \underline{x}^TQ_0\underline{x})^{-1}Q_0\underline{x}\underline{x}^TQ_0]^{-1}\underline{h} \tag{3.68}$$

and further as (cf. Schaffrin [1999])

$$(N + Q_0)\underline{\hat{x}}_{BLSUE} = \underline{h} + \alpha Q_0\underline{x}, \tag{3.69}$$

with

$$\alpha = \frac{\underline{x}^TQ_0(N + Q_0)^{-1}\underline{h}}{1 + \underline{x}^TQ_0(N + Q_0)^{-1}N\underline{x}}. \tag{3.70}$$

Equivalent to Schaffrin's Robust Collocation theory, which uses the mixed-linear model as a functional model ([Schaffrin, 1985, 1986, 1987], [Middel and Schaffrin, 1987]), one can replace the vector $\underline{x}$ by the vector of prior information $\underline{\hat{x}}_P$ and Q_0 by the normal matrix N_P. The estimator then reads

$$(N + N_P)\underline{\hat{x}}_{RC} = \underline{h} + \alpha N_P\underline{\hat{x}}_P \tag{3.71}$$

with

$$\alpha = \frac{\underline{\hat{x}}_P^TN_P(N + N_P)^{-1}\underline{h}}{1 + \underline{\hat{x}}_P^TN_P(N + N_P)^{-1}N\underline{\hat{x}}_P}. \tag{3.72}$$

Note that the parameter α is dependent on the values of the right-hand-side vectors $N_P\underline{\hat{x}}_P$ and $\underline{h}$ and therefore is not invariant to translations. As the BLSUE is biased and not translation invariant, we suggest not to use the BLSUE in global gravity field modelling. If the solution is highly sensitive to errors in the prior information, one should either change the stochastic model of the prior information or augment the functional model to account for inconsistencies between the prior information and the new data.

3.5 Summary

The linear relation between the observations and the parameters (to be estimated) are captured in the design matrix, as part of the functional model. Any systematic effects in a data set should be accounted for in this functional model. If the functional model is not properly defined, the estimated parameters will absorb these systematic effects. Moreover, in a combination of different data sets, the unmodelled systematic effects will cause inconsistencies between the data sets.

The functional model should be augmented by extra (local) parameters to account for these systematic effects. The residual square sum will always decrease with the addition of extra parameters. However, one should first test the significance of the parameters before implementing them. The addition of a data set requires that the data set be consistent with the other data sets. Tests have been derived to test the data set for unbiasedness, or test for consistencies between the data sets.

The chapter ends with a discussion of the Best Linear Softly Unbiased Estimator, which weakens the constraint of unbiasedness in the combination of the data sets, and robustify the addition of the extra data set.

4

Stochastic model validation

In the previous chapter, we tried to eliminate any systematic errors from the vector of residuals by augmentation of the functional model. The remaining vector of residuals between the observations and the estimated signal is assumed to be stochastic with random noise. If this noise is normally distributed with zero mean, we can express its stochastic properties by its dispersion, the variance-covariance (VC) matrix.

A proper choice of the VC-matrix is necessary for the estimation of the vector of unknowns in a weighted least-square sense and the quality description of this estimation. Moreover, we will show later that this matrix is crucial in the detection and treatment of outliers.

To estimate the elements of the VC-matrix, we have to make use of the redundancy in the observations. However, this redundancy is not sufficient to estimate all elements separately. The matrix is therefore written as a linear combination of different *cofactor* matrices with its scaling factors (*variance components*) to be estimated.

Several approaches have been proposed in the past to estimate these variance components. Many of these estimators converge, under normality, to the same estimate. The estimators are unbiased, translation invariant, have minimum variance, maximum likelihood and are optimal in a weighted least-squares sense. In the early seventies, C.R. Rao [Rao, 1971a] published his paper on the MINQUE-theory, which leads to the estimator with the above-mentioned properties (under normality). The estimator is equal to the estimator of Helmert [1924] if and only if the observations are normally distributed and the observation groups are uncorrelated with one variance component to be estimated for each observation group.

Rao [1971b] showed that the MINQUE (under normality) is of minimum variance and Sjöberg [1983] estimated the VC-matrix of the estimated variance components. In Teunissen [1988] and recently in Amiri-Simkooei [2007], the equations were derived following a weighted least-squares approach. Maximization of the likelihood, under consideration of the degrees of freedom involved in the estimation of the vector of un-

knowns, has been introduced by Harville [1977] and later modified by Koch [1986, 1990].

A fast algorithm, which in general converge to the same estimate, was first developed by Horn et al. [1975] under the assumption of uncorrelated observations. This *Iterative Restricted Maximum Likelihood Estimator* was further generalized by Förstner [1979] and Koch [1986].

Alternative methods, which converge to a different estimate, are the Bayesian estimate ([Koch, 1990], [Ou, 1991, 1993]), Lerch's subset solution method [Lerch, 1989, 1991] and the *Iterative Maximum Likelihood Estimator* (IMLE) [Yuan, 1991]. In Sjöberg [1984] a non-negative alternative to MINQUE is suggested. However, negative variance components indicate either a low redundancy, an improperly designed variance component model or badly chosen a-priori variance components [Amiri-Simkooei, 2007]. Therefore, these estimators will not be addressed. In Schwintzer [1990], Koch and Kusche [2002] and Xu et al. [2006], VCE is used to estimate the regularization parameter in an ill-posed problem.

We will start this chapter with the derivation of the properties of the MINQUE estimator. We then re-write and simplify these equations under the consideration of uncorrelated observation groups. The derivation of the alternative algorithms is the subject of section 4.6. Further time-reducing operations, using Monte Carlo simulations, will be dealt with in chapter 5.

4.1 The stochastic model

The stochastic model represents the stochastic properties of the vector of residuals $\underline{e}$ (assuming a Gauss-Markov functional model). This vector contains the misfit between the observation vector $\underline{y}$ and the model $A\underline{x}$. The vector of residuals is assumed to be unbiased, i.e. $E\{\underline{e}\} = \underline{0}$. Any systematic effect should be either removed from the observations or added as nuisance parameters to the functional model. However, in satellite gravity modelling, it is, in general, not possible to exclude all these errors from the residual vector. Fixed effects, such as errors in the tidal or atmospheric models, as well as truncation errors, remain in the vector of residuals $\underline{e}$.

Along with these systematic effects, the vector $\underline{e}$ consists of stochastic errors, which are caused by many sources. Summation of different elementary errors, each with an arbitrary probability function, generally results in an error vector with a multivariate normal distribution (according to the *central limit theorem*; see, e.g., Cramér [1946]). Hence, the error vector can be written as a linear combination of $\bar{q}$ elementary errors (constituents):

$$\underline{e} = U_1\underline{\epsilon}_1 + ... + U_{\bar{q}}\underline{\epsilon}_{\bar{q}} = \sum_{k=1}^{\bar{q}} U_k\underline{\epsilon}_k \tag{4.1}$$

with

U_k = $m \times a_k$ transformation matrix
$\underline{\epsilon}_k$ = $a_k \times 1$ vector of stochastic errors of group k.

These elementary errors are assumed to be unbiased, i.e.

$$\begin{array}{lcll} E\{\underline{\epsilon}_k\} & = & 0 & \text{for all } k = 1, ..., \bar{q} \\ D\{\underline{\epsilon}_k\} & = & \bar{\sigma}_k^2 Q_{\epsilon_k} & \text{for all } k = 1, ..., \bar{q} \\ C\{\underline{\epsilon}_k, \underline{\epsilon}_l\} & = & \bar{\sigma}_{kl} Q_{\epsilon_k \epsilon_l} & \text{for all } k, l = 1, ..., \bar{q}, k \neq l \end{array} \tag{4.2}$$

with

$$\begin{array}{ll} \bar{\sigma}_k^2 & = \text{variance component} \\ \bar{\sigma}_{kl} & = \text{covariance component.} \end{array}$$

Note that the dimension of the (co)variance components equals the square of the dimension of the observations and the Q_{ϵ_k} and $Q_{\epsilon_k \epsilon_l}$ matrices are dimensionless. As the covariance component $\bar{\sigma}_{kl}$ equals the covariance component $\bar{\sigma}_{lk}$, the total number of variance and covariance components is equal to $q = \bar{q}(\bar{q}+1)/2$. Hence, the variance-covariance matrix can be split into q matrices:

$$Q_y = \sum_{i=1}^{q} \gamma_i Q_i \tag{4.3}$$

where

$$\begin{array}{llllll} Q_i = U_k Q_{\epsilon_k} U_k^T & ; & \gamma_i = \bar{\sigma}_k^2 & ; & \text{for all} & k = 1, ..., \bar{q}; \\ Q_i = U_k Q_{\epsilon_k \epsilon_l} U_l^T + U_l Q_{\epsilon_l \epsilon_k} U_k^T & ; & \gamma_i = \bar{\sigma}_{kl} & ; & \text{for all} & k = 1, ..., \bar{q}, \quad l < k. \end{array} \tag{4.4}$$

For simplicity, we shall call all γ_i elements variance components from now on, as in practice the vast majority of the (co)variance components are in fact variance components. The matrices Q_{ϵ_k} and $Q_{\epsilon_k \epsilon_l}$ are, however, generally not known exactly. The choice of the *cofactor* matrices Q_i is based on prior assumptions and to a certain extent arbitrary. Note that different linear combinations of the right-hand side of Eq. (4.3) may result in the same matrix Q_y. This means, the choice of a model for the variance-covariance matrix Q_y is non-unique. Several variance component estimators will now be introduced.

4.2 Minimum Norm Quadratic Unbiased Estimator (MINQUE)

One of the best known estimators for the variance components (in a linear stochastic model) is the Minimum Norm Quadratic Unbiased Estimator (MINQUE), defined by C.R. Rao in the early seventies, see, e.g., Rao [1971a, 1973] and Rao and Mitra [1971]. We will derive the equations for the MINQUE by looking at the different properties of the estimator, which are unbiasedness, invariance and a minimum Euclidean norm.

Quadratic Estimator

If a linear combination of variance components, i.e., $\underline{l}^T\underline{\gamma}$, is estimated by a quadratic form of the observations:

$$\underline{y}^T M \underline{y},$$

where M is an arbitrary, symmetric $m \times m$ matrix, $\underline{y}^T M \underline{y}$ is said to be a Quadratic Estimator (QE) of $\underline{l}^T\underline{\gamma}$. The expectation of the Quadratic Estimator reads, using the Gauss-Markov model:

$$\begin{aligned} E\{\underline{y}^T M \underline{y}\} &= E\{(A\underline{x} + \underline{e})^T M (A\underline{x} + \underline{e})\} \\ &= \underline{x}^T A^T M A \underline{x} + 2\underline{x}^T A^T M E\{\underline{e}\} + E\{\underline{e}^T M \underline{e}\} \\ &= \underline{x}^T A^T M A \underline{x} + \text{tr}(MQ_y). \end{aligned} \tag{4.5}$$

Invariance with respect to a datum shift

The estimator should be invariant for any shift in datum, such as $\underline{x}_d := \underline{x} - \underline{x}_0$; e.g., the reduction of a reference gravity model from the measurements $\underline{y}$ and the vector of unknowns $\underline{x}$:

$$\underline{y}_d = A\underline{x}_d + \underline{e} = \underline{y} - A\underline{x}_0. \tag{4.6}$$

The estimated variance components should not depend on a particular choice of datum or reference model. Hence, the estimator $\underline{y}_d^T M \underline{y}_d^T$ must be equal to $\underline{y}^T M \underline{y}$ for any datum shift $\underline{x}_0$:

$$\begin{aligned} \underline{y}_d^T M \underline{y}_d &= (\underline{y} - A\underline{x}_0)^T M (\underline{y} - A\underline{x}_0) \\ &= \underline{y}^T M \underline{y} - 2\underline{x}_0^T A^T M \underline{y} + \underline{x}_0^T A^T M A \underline{x}_0. \end{aligned} \tag{4.7}$$

This equals $\underline{y}^T M \underline{y}$ if

$$MA = O. \tag{4.8}$$

Note that MA is a $m \times n$ matrix, with m the number of observations and n the number of unknown parameters. When these $m \cdot n$ equations hold, the estimator is invariant with respect to the vector $\underline{x}$. Equation (4.5) then reduces to:

$$E\{\underline{y}^T M \underline{y}\} = \text{tr}(MQ_y). \tag{4.9}$$

The matrix M can be formally split into q $m \times m$ matrices for which Eq. (4.8) must be valid, with q the number of variance components to be estimated:

$$M = \sum_{j=1}^{q} \lambda_j M_j. \tag{4.10}$$

Introducing the $q \times 1$ vector $\underline{u}$, with its j-th component

$$u_j = \underline{y}^T M_j \underline{y} \tag{4.11}$$

gives

$$\begin{aligned} E\{u_j\} &= E\{\underline{y}^T M_j \underline{y}\} = \text{tr}(M_j Q_y) \\ &= \sum_{i=1}^{q} \gamma_i \text{tr}(M_j Q_i) = \sum_{i=1}^{q} \gamma_i S_{ij}. \end{aligned} \tag{4.12}$$

For an Invariant Quadratic Estimator (IQE), which is split according to Eq. (4.10), the following relation holds:

$$E\{\underline{u}\} = S\underline{\gamma} \tag{4.13}$$

with

$$\begin{aligned} S_{ij} &= \text{tr}(M_j Q_i) \\ u_j &= \underline{y}^T M_j \underline{y}. \end{aligned}$$

The choice of the λ_j coefficients and the M_j matrices is still arbitrary. The $\underline{\lambda}$ vector will be defined when requiring unbiasedness. The M_j matrices will be defined when optimizing the variance components estimation, e.g., maximum likelihood, minimum variance, etc.

Unbiasedness of the estimator

The linear combination of the variance components, $\underline{l}^T \underline{\gamma}$, is estimated unbiasedly by $\underline{y}^T M \underline{y}$ if and only if

$$E\{\underline{y}^T M \underline{y}\} = \underline{l}^T \underline{\gamma}. \tag{4.14}$$

This is true if [Rao, 1973]

$$\text{tr}(MQ_i) = l_i, \tag{4.15}$$

as

$$E\{\underline{y}^T M \underline{y}\} = \text{tr}(MQ_y) = \sum_{i=1}^{q} \gamma_i \text{tr}(MQ_i) = \sum_{i=1}^{q} \gamma_i l_i. \tag{4.16}$$

Equation (4.15) can be rewritten as

$$l_i = \text{tr}(MQ_i) = \sum_{j=1}^{p} \lambda_j \text{tr}(M_j Q_i) = \sum_{j=1}^{q} \lambda_j S_{ij}. \tag{4.17}$$

Hence, an Invariant Quadratic Unbiased Estimator (IQUE) is defined as an IQE with the additional property that the λ_i defined in (4.10) fulfill

$$S\underline{\lambda} = \underline{l}. \tag{4.18}$$

The definition of an IQUE does not include any assumption on the distribution of the observations $\underline{y}$ apart from (3.1). The IQUE provides an unbiased, quadratic estimator of a linear combination of the variance components and is invariant with respect to a translation in the $\underline{x}$-parameter space. The choice of the M_j-matrices is still arbitrary. The determination of the 'best' set of M_j matrices will now be addressed.

Minimum Euclidean norm

The MINQUE is defined as the IQUE with minimum norm [Rao, 1973]. Although many matrix norms exist, in MINQUE one considers the Euclidean norm, i.e., the sum of the squares of all the elements of a matrix. This is equivalent to the minimization of [Rao, 1973]

$$\text{tr}(MQ_yMQ_y). \tag{4.19}$$

Hence, one needs to minimize $\text{tr}(MQ_yMQ_y)$ subject to the constraints (4.8) and (4.15). This is equal to minimizing the Lagrange function $\Psi(M)$ [Koch, 1999]:

$$\Psi(M) = \text{tr}(MQ_yMQ_y) - 2\text{tr}(MA\Lambda^T) - 2\sum_{j=1}^{q}\lambda_j\text{tr}(MQ_j - p_j) \tag{4.20}$$

where -2Λ denotes the $m \times n$ matrix of Lagrange multipliers for the constraints (4.8) and $-2\lambda_j$ the q Lagrange multipliers for the constraints (4.15). The Lagrange function can be minimized by setting its derivative equal to zero. Using the properties (see Eq. (B.1) and (B.2))

$$\frac{\partial\text{tr}(AB)}{\partial A} = B^T \tag{4.21}$$

and

$$\frac{\partial\text{tr}(ABAB)}{\partial A} = 2(BAB)^T, \tag{4.22}$$

this derivative becomes

$$\frac{\partial\Psi}{\partial M} = 2Q_yMQ_y - 2\Lambda A^T - 2\sum_{j=1}^{q}\lambda_jQ_j. \tag{4.23}$$

Setting this equal to zero results in

$$Q_y\hat{M}Q_y = \Lambda A^T + \sum_{j=1}^{q}\lambda_jQ_j. \tag{4.24}$$

To get rid of the Q_y-matrices on the left-hand-side, one has to multiply this equation from the left and right by Q_y^{-1}. The first term on the right-hand-side, ΛA^T can be eliminated by a post-multiplication with the projection matrix $(P_A^\perp)^T$. The symmetric matrix R, also known as the redundancy or Rao's matrix does both:

$$R := Q_y^{-1}P_A^\perp = (P_A^\perp)^TQ_y^{-1} = Q_y^{-1} - Q_y^{-1}A(A^TQ_y^{-1}A)^{-1}A^TQ_y^{-1}. \tag{4.25}$$

The minimum norm is attained if [Rao, 1973]

$$\hat{M} = \sum_{j=1}^{q}\lambda_jRQ_jR. \tag{4.26}$$

Hence,

$$\hat{M}_j = RQ_jR. \tag{4.27}$$

When substituting Eq. (4.27) into Eq. (4.13), we find the system of conditions for a MINQUE:

$$E\{\underline{u}\} = S\underline{\gamma} \tag{4.28}$$

with

$$\begin{aligned} S_{ij} &= \text{tr}(RQ_iRQ_j) \\ u_j &= \underline{y}^T RQ_jR\underline{y}. \end{aligned}$$

Solving this system for $\underline{\gamma}$ is an iterative procedure, as the computation of Rao's matrix R involves the unknown variance components. After defining $\bar{\gamma}_i$ as an approximate value for the variance component γ_i, an improved estimate $\hat{\gamma}_i$ can be computed as

$$\underline{\hat{\gamma}} = S^{-1}(\underline{\bar{\gamma}})\underline{u}(\underline{\bar{\gamma}}). \tag{4.29}$$

This estimate will then be used as the approximate value for the variance components in the next iteration. The algorithm is repeated until convergence. As the unbiased estimator of Eq. (4.28) requires the stochastic model to be known (and true), the iterations of MINQUE, using an a-priori estimate of the stochastic model, will not be unbiased. However, we will assume that this bias will be reduced within each iteration step and can almost be neglected at convergence.

4.3 Properties of the MINQUE under normality

Several estimators have been proposed in the past to estimate the variance components. However, some of these estimators produce the same equations as MINQUE and one should look at them as proofs for certain properties of the MINQUE. Three estimators will be dealt with, each of them assuming the data to be normally distributed. It has already been shown in the previous section that in general a MINQUE

- is unbiased,
- is invariant to a datum translation, and
- has minimum norm.

Under normality, we will prove in this section that a MINQUE also

- has minimum variance (MIVQUE, BIQUE),
- can be derived in a least-squares sense (LSVCE), and
- converges to the (restricted) maximum likelihood (REML) estimator.

Moreover, a quality description of the estimated variance components in terms of a variance-covariance matrix is available under normality.

4.3.1 Best Invariant Quadratic Unbiased Estimator (BIQUE)

Rao [1971b] derived equations for the Minimum Variance Quadratic Unbiased Estimator (MIVQUE) and found that this estimator equals the MINQUE if the observations are normally distributed. This was also proved by Sjöberg [1983], who derived the equations for MINQUE under normality and assumed the general linear adjustment model (condition adjustment with unknowns) as the functional model, of which the Gauss-Markov functional model is a special case. This estimator is addressed as the *Best Invariant Quadratic Unbiased Estimator (BIQUE)*. The variance of the Quadratic Estimator $\underline{y}^T M \underline{y}$, in which $\underline{y}$ is a vector of normally distributed data, reads [Rao and Mitra, 1971]

$$D\{\underline{y}^T M \underline{y}\} = 4\underline{x}^T A^T M Q_y M A \underline{x} + 2\text{tr}(M Q_y M Q_y). \tag{4.30}$$

Under invariance ($MA = 0$) this reduces to

$$D\{\underline{y}^T M \underline{y}\} = 2\text{tr}(M Q_y M Q_y). \tag{4.31}$$

Minimizing the variance thus is equivalent to minimizing the norm, as was done in the derivation of the MINQUE, see Eq. (4.19). The BIQUE is, therefore, equal to the MINQUE under the assumption of normally distributed observations. The covariance between two linear combinations of estimated variance components can be derived as follows:

$$\begin{aligned} C\{\underline{p}^T\hat{\underline{\gamma}}, \underline{q}^T\hat{\underline{\gamma}}\} &= C\{\underline{y}^T\hat{M}\underline{y}, \underline{y}^T\hat{N}\underline{y}\} = 2\text{tr}(\hat{M}Q_y\hat{N}Q_y) \\ &= 2\text{tr}([\sum_{i=1}^{q}\lambda_i R Q_i R]Q_y[\sum_{j=1}^{q}\mu_j R Q_j R]Q_y) \\ &= 2\text{tr}(\sum_{i=1}^{q}\lambda_i R Q_i[\sum_{j=1}^{q}\mu_j R Q_j]) \\ &= 2\sum_{i=1}^{q}\sum_{j=1}^{q}\lambda_i\mu_j\text{tr}(R Q_i R Q_j) \\ &= 2\underline{\lambda}^T S\underline{\mu} = 2\underline{\lambda}^T S S^{-1} S\underline{\mu} = 2\underline{p}^T S^{-1}\underline{q}. \end{aligned} \tag{4.32}$$

Hence, the VC-matrix of the estimator $\hat{\underline{\gamma}}$ reads [Sjöberg, 1983]

$$Q_{\hat{\gamma}} = 2S^{-1}. \tag{4.33}$$

4.3.2 Least-Squares Variance Component Estimation (LSVCE)

In Teunissen [1988], the variance components are written as a vector of unknowns in a least-squares setting, see also Pukelsheim [1976]. New applications and a more elaborate study on this subject can be found in Amiri-Simkooei [2007], Amiri-Simkooei et al. [2007] and Teunissen and Amiri-Simkooei [2008].

Let B be a $m \times (m-n)$ basis matrix of the null space of A^T, i.e., the columns of B span the null space of A^T, $\mathcal{N}(A^T)$,and $A^T B = O_{n\times(m-n)}$ holds. The $(m-n)\times 1$

vector of misclosures $\underline{v}$ is defined as

$$\underline{v} := B^T \underline{y} = B^T \underline{e}, \tag{4.34}$$

with the stochastic model

$$E\{\underline{v}\} = \underline{0} \quad ; \quad D\{\underline{v}\} = B^T Q_y B = Q_v. \tag{4.35}$$

Basically, one makes use of the remaining redundancy $r = m - n$ after estimating the least-squares estimate $\underline{\hat{x}}$ to estimate the variance component vector $\underline{\gamma}$. The $\underline{v}$ and $\underline{\hat{x}}$ vectors are therefore uncorrelated. As the misclosure vector has zero mean, it follows that

$$E\{\underline{v}\underline{v}^T\} = B^T Q_y B = \sum_{i=1}^{p} B^T Q_i B \gamma_i. \tag{4.36}$$

The *observation matrix* $\underline{v}\underline{v}^T$ has $r \cdot (r+1)/2$ distinct elements. We will use the vh-operator, which stacks the columns of the lower part of a symmetric matrix below each other. One should not use the vec-operator, as this would result in a singular variance matrix of the observation vector $\text{vec}\{\underline{v}\underline{v}^T\}$ [Teunissen and Amiri-Simkooei, 2008]. Applying the vh-operator on the observation equation results in a general linear Gauss-Markov model:

$$\underline{y}_{vh} = A_{vh}\underline{\gamma} + \underline{e}_{vh} \tag{4.37}$$

with y_{vh} the new $r \cdot (r+1)/2 \times 1$ observation vector, defined as

$$\underline{y}_{vh} := \text{vh}\{\underline{v}\underline{v}^T\}, \tag{4.38}$$

and the new $r \cdot (r+1)/2 \times q$ design matrix A_{vh}, defined as

$$A_{vh} := [\text{vh}\{B^T Q_1 B\}, \ldots, \text{vh}\{B^T Q_q B\}], \tag{4.39}$$

and the stochastic model

$$E\{\underline{e}_{vh}\} = \underline{0} \quad ; \quad D\{\underline{e}_{vh}\} = Q_{vh}. \tag{4.40}$$

Hence, one could only solve for a variance component model with a maximum of $r \cdot (r+1)/2$ variance components. The BLUE of $\underline{\gamma}$, which is a Best Quadratic Unbiased Estimator with respect to the observation vector $\underline{y}$, therefore reads

$$N\underline{\hat{\gamma}} = \underline{h}, \tag{4.41}$$

with

$$\begin{aligned} N &= A_{vh}^T Q_{vh}^{-1} A_{vh} \\ \underline{h} &= A_{vh}^T Q_{vh}^{-1} \underline{y}_{vh}. \end{aligned}$$

If we make use of the *duplication matrix* D with the property $D \cdot \text{vh}\{M\} = \text{vec}\{M\}$, we can state that [Amiri-Simkooei, 2007]

$$Q_{vh}^{-1} = \frac{1}{2} D^T (Q_v^{-1} \otimes Q_v^{-1}) D. \tag{4.42}$$

The elements of the matrix N can now be written as

$$\begin{aligned} N_{ij} &= \frac{1}{2}(\text{vec}\{B^T Q_i B\})^T (Q_v^{-1} \otimes Q_v^{-1}) \text{vec}\{B^T Q_j B\} \\ &= \frac{1}{2}\text{tr}(Q_v^{-1} B^T Q_j B Q_v^{-1} B^T Q_i B) \\ &= \frac{1}{2}\text{tr}(R Q_i R Q_j) \\ &= \frac{1}{2} S_{ij}, \end{aligned} \tag{4.43}$$

where use is made of [Teunissen et al., 2005]

$$A(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} + Q_y B (B^T Q_y B)^{-1} B^T = I_m. \tag{4.44}$$

and of the identity of Eq. (B.11). In a similar way, we can compute the elements of $\underline{h}$:

$$\begin{aligned} h_i &= \frac{1}{2}(\text{vec}\{B^T Q_i B\})^T (Q_v^{-1} \otimes Q_v^{-1}) \text{vec}\{\underline{v}\underline{v}^T\} \\ &= \frac{1}{2}\underline{v}^T Q_v^{-1} B^T Q_i B Q_v \underline{v} \\ &= \frac{1}{2}\underline{y}^T R Q_i R \underline{y} \\ &= \frac{1}{2} u_i. \end{aligned} \tag{4.45}$$

Note that

$$\underline{\hat{\gamma}} = N^{-1}\underline{h} = S^{-1}\underline{u} \quad ; \quad Q_{\hat{\gamma}} = N^{-1} = 2S^{-1}. \tag{4.46}$$

Hence, the LSVCE is, under normality, identical to the BIQUE and the MINQUE.

4.3.3 Restricted Maximum Likelihood Estimation (REML)

We will now show that the equations to compute the MINQUE also lead to maximum likelihood (at convergence) if we can assume normally distributed observations. The method has been proposed by Patterson and Thompson [1971] and Harville [1977] and was later modified by Koch [1986,1990]. The probability density function for a multivariate normal distribution reads

$$p(\underline{y}|\underline{x}, \underline{\gamma}) = \frac{1}{(2\pi)^{m/2}(\det Q_y)^{1/2}} \exp[-\frac{1}{2}(\underline{y} - A\underline{x})^T Q_y^{-1}(\underline{y} - A\underline{x})]. \tag{4.47}$$

The probability density function is dependent on the variance-covariance matrix Q_y (and consequently on the variance components vector $\underline{\gamma}$) and is dependent on the vector

of unknowns $\underline{x}$. If we try to find the maximum of this function by varying one of the two vectors ($\underline{\gamma}$ or $\underline{x}$) and *fixing* the other vector, we will get a Maximum Likelihood Estimator (MLE). This estimator is however biased and has no minimum variance. The MLE will be discussed in section 4.6.5.

In this section we will consider the *Restricted* Maximum Likelihood Estimator, in which the vector of unknown parameters $\underline{x}$ will be seen as a vector of *free* parameters and integrated out of the likelihood function. So it maximizes that part of the likelihood function, which is invariant to the vector of unknown parameters $\underline{x}$. A non-informative prior for this vector is introduced, which results in a constant value for the conditional density function $p(\underline{x}|\underline{y})$. The conditional density function $p(\underline{y}|\underline{\gamma})$ can be computed by

$$p(\underline{y}|\underline{\gamma}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\underline{y}|\underline{x},\underline{\gamma}) p(\underline{x}|\underline{y}) \mathrm{d}\underline{x}. \tag{4.48}$$

In Koch [1990] it was found that this equation (under the assumption of normally distributed data) can be simplified to

$$p(\underline{y}|\underline{\gamma}) \propto \frac{1}{(\mathrm{det}Q_y \mathrm{det}(A^T Q_y^{-1} A))^{1/2}} \exp(-\frac{1}{2}\underline{y}^T R \underline{y}), \tag{4.49}$$

where $\propto$ denotes proportionality and R is Rao's matrix (Eq. 4.25). We make use of proportionality relations whenever possible because this allows us to disregard unnecessary constants. The REML can now be found by maximizing this probability density function by varying the variance components. This maximization can be simplified when using the natural logarithm of the likelihood function:

$$\ln p(\underline{y}|\underline{\gamma}) \propto C - \ln \mathrm{det}Q_y - \ln \mathrm{det}(A^T Q_y^{-1} A) - \underline{y}^T R \underline{y}, \tag{4.50}$$

where C is some constant. This equation is differentiated with respect to γ_i. Using the properties (Eq. B.4 and Eq. B.5) of a symmetric matrix, the derivative becomes

$$\begin{aligned} \frac{\partial \ln p(\underline{y}|\underline{\gamma})}{\partial \gamma_i} &\propto -\frac{\partial \ln \mathrm{det}Q_y}{\partial \gamma_i} - \frac{\partial \ln \mathrm{det}(A^T Q_y^{-1} A)}{\partial \gamma_i} - \frac{\partial \underline{y}^T R \underline{y}}{\partial \gamma_i} \\ &= -\mathrm{tr}(Q_y^{-1} Q_i) + tr[(A^T Q_y^{-1} A)^{-1} A^T Q_y^{-1} Q_i Q_y^{-1} A] - \underline{y}^T \frac{\partial R}{\partial \gamma_i} \underline{y} \\ &= -\mathrm{tr}(RQ_i) + \underline{y}^T R Q_i R \underline{y}; \end{aligned} \tag{4.51}$$

see also Koch [1986,1990]. Setting this equation equal to zero results in a set of conditions, which determine the REML of the variance components [Koch, 1986]. Condition for the Restricted Maximum Likelihood Estimator:

$$\underline{t} - \underline{u} = \underline{0} \tag{4.52}$$

where

$$\begin{aligned} t_i &= \mathrm{tr}(RQ_i) \\ u_i &= \underline{y}^T R Q_i R \underline{y}. \end{aligned}$$

Newton-Raphson iteration

The q conditions of the REML of the variance components, as are stated in Eq. (4.52), can be solved with a Newton-Raphson (NR) iteration. First, we introduce the residual vector $\underline{d}(\underline{\gamma})$:

$$\underline{d}(\underline{\gamma}) = \underline{t}(\underline{\gamma}) - \underline{u}(\underline{\gamma}). \tag{4.53}$$

Using the vector $\bar{\underline{\gamma}}$ as an approximation of the solution of Eq. (4.52), NR iteration can be used to predict a better estimate $\hat{\underline{\gamma}}$ of the variance components:

$$\underline{d}(\underline{\gamma}) = \underline{d}(\bar{\underline{\gamma}}) + \left.\frac{\partial \underline{d}(\underline{\gamma})}{\partial \underline{\gamma}}\right|_{\underline{\gamma}=\bar{\underline{\gamma}}} (\hat{\underline{\gamma}} - \bar{\underline{\gamma}}) = \underline{0}. \tag{4.54}$$

Rewriting Eq. (4.54) gives the equation to compute $\hat{\underline{\gamma}}$:

$$\hat{\underline{\gamma}} = \bar{\underline{\gamma}} - H^{-1}(\bar{\underline{\gamma}})\underline{d}(\bar{\underline{\gamma}}), \tag{4.55}$$

where H is the Hessian matrix [Koch, 1986]:

$$H_{ij} = \left.\frac{\partial d_i(\underline{\gamma})}{\partial \gamma_j}\right|_{\underline{\gamma}=\bar{\underline{\gamma}}} = 2\underline{y}^T RQ_jRQ_iR\underline{y} - \text{tr}(RQ_jRQ_i) = K_{ij} - S_{ij}. \tag{4.56}$$

Rewriting Eq. (4.55) results in an expression to compute the REML using NR iteration:

$$\begin{array}{rl} H\hat{\underline{\gamma}} & = \underline{u}(\bar{\gamma}) + H\bar{\gamma} - \underline{t}(\bar{\gamma}) \\ & = \underline{u}(\bar{\gamma}) + K\bar{\gamma} - S\bar{\gamma} - \underline{t}(\bar{\gamma}) \end{array} \tag{4.57}$$

This system of equations can be simplified using

$$(S\bar{\underline{\gamma}})_i = \sum_{j=1}^{q} S_{ij}\bar{\gamma}_j = \sum_{j=1}^{q} \text{tr}(RQ_jRQ_i)\bar{\gamma}_j = \text{tr}(RQ_yRQ_i) = \text{tr}(RQ_i) = t_i. \tag{4.58}$$

and

$$(K\bar{\underline{\gamma}})_i = \sum_{j=1}^{q} K_{ij}\bar{\gamma}_j = \sum_{j=1}^{q} 2\underline{y}^T RQ_jRQ_iR\underline{y}\bar{\gamma}_j = 2\underline{y}^T RQ_iR\underline{y} = 2u_i. \tag{4.59}$$

Hence, the system of equations to compute an improved estimate of the REML using NR iteration reads

$$H\hat{\underline{\gamma}} = 3\underline{u} - 2\underline{t} \tag{4.60}$$

with

$$\begin{array}{rcl} H_{ij} & = & 2\underline{y}^T RQ_jRQ_iR\underline{y} - \text{tr}(RQ_iRQ_j) \\ u_j & = & \underline{y}^T RQ_jR\underline{y} \\ t_j & = & \text{tr}(RQ_j). \end{array}$$

We will abbreviate the REML using Newton-Raphson iterations as *REML-nr*.

Fisher scoring

The Hessian matrix depends on the observations $\underline{y}$. The computational effort to compute the REML can be reduced by replacing the Hessian matrix by its expectation $E\{H\}$ (see, e.g., Koch [1986] and Searle et al. [1992]):

$$\begin{aligned} E\{H_{ij}\} &= E\{2\underline{y}^T RQ_jRQ_iR\underline{y}\} - \text{tr}(RQ_jRQ_i) \\ &= 2\text{tr}(RQ_jRQ_iRE\{\underline{y}\underline{y}^T\}) - \text{tr}(RQ_jRQ_i) \\ &= 2\text{tr}(RQ_jRQ_iRQ_y) - \text{tr}(RQ_jRQ_i) \\ &= 2\text{tr}(RQ_yRQ_jRQ_i) - \text{tr}(RQ_jRQ_i) \\ &= 2\text{tr}(RQ_jRQ_i) - \text{tr}(RQ_jRQ_i) \\ &= \text{tr}(RQ_jRQ_i) \\ &= S_{ij}. \end{aligned} \tag{4.61}$$

This is called Fisher scoring. The iteration formula for the Fisher scoring thus reads:

$$\hat{\underline{\gamma}} = \bar{\underline{\gamma}} - S^{-1}(\bar{\underline{\gamma}})\underline{d}(\bar{\underline{\gamma}}) \tag{4.62}$$

or

$$S(\hat{\underline{\gamma}} - \bar{\underline{\gamma}}) = \underline{u}(\bar{\underline{\gamma}}) - \underline{t}(\bar{\underline{\gamma}}) \tag{4.63}$$

with $\underline{u}$ and $\underline{t}$ according to Eq. (4.52). This system of equations can be simplified using Eq. (4.58). Hence, the system of equations to compute an improved estimate of the REML using Fisher scoring (REML-fs) reads

$$S\hat{\underline{\gamma}} = \underline{u} \tag{4.64}$$

with

$$\begin{aligned} S_{ij} &= \text{tr}(RQ_iRQ_j) \\ u_j &= \underline{y}^T RQ_jR\underline{y}, \end{aligned}$$

which proves that MINQUE, under normality, converges to the Restricted Maximum Likelihood (REML). However, the iterations of MINQUE will in general not have maximum likelihood.

4.3.4 Other properties

As we can write the vector of variance components as a vector of unknowns in the Gauss-Markov model and estimate them in a least-squares sense, we can test the estimation of the variance components for significance and derive the equations for linearization and non-negative estimation [Amiri-Simkooei, 2007]. In this section, we will shortly address these topics.

Estimability

Note that a linear independence of the cofactor matrices is a necessary but not a sufficient condition for the estimability of the variance components. The columns of $A_{vh} := [\text{vh}\{B^T Q_1 B\}, \ldots, \text{vh}\{B^T Q_q B\}]$ should be independent of each other, i.e. the matrix A_{vh} should have full rank. One of the conditions for this is that the number of variance components to be estimated is not larger than $r \cdot (r+1)/2$, with $r = m - n$.

Test for significance

If a variance component can be estimated, we should test if this variance component is significant. Therefore, we need to test

$$H_0 : \gamma_i = 0 \quad ; \quad H_A : \gamma_i \neq 0. \tag{4.65}$$

In Amiri-Simkooei [2007], it was found that this can be tested by the v-test statistic:

$$v_i = \frac{\underline{\hat{e}}^T Q_y^{-1} \sum_{j=1}^{q} [Q_{\hat{\gamma}}]_{ij} Q_j Q_y^{-1} \underline{\hat{e}}}{2\sqrt{[Q_{\hat{\gamma}}]_{ii}}}, \tag{4.66}$$

which has a non-trivial distribution, i.e., a sum of chi-squared distributions; see Amiri-Simkooei [2007].

Prior information

If we have reliable, independent information on the vector of variance components, i.e., $\underline{y}_P = H\underline{\gamma}_P$, with a reliable estimate of its stochastic properties, i.e., $D\{\underline{y}_P\} = HQ_{\gamma_P}H^T$, then the combined normal equations will read

$$\left[\frac{1}{2}S + H^T \left(HQ_{\gamma_P}H^T\right)^{-1} H\right] \underline{\hat{\gamma}} = \frac{1}{2}\underline{u} + H^T \left(HQ_{\gamma_P}H^T\right)^{-1} \underline{y}_P. \tag{4.67}$$

Linearization

Up till now we have assumed the stochastic model to be a linear function of several cofactor matrices, with the variance components $\underline{\gamma}$ the linear coefficients. Now, we will assume the non-linear case:

$$Q_y = Q(\underline{\gamma}). \tag{4.68}$$

Using the equations for the derivation of the LSVCE, the linearized functional model reads

$$\Delta\underline{y}_{vh} = A_{vh}\Delta\underline{\gamma} + \underline{e}_{vh}, \tag{4.69}$$

where

$$\begin{aligned} \Delta\underline{y}_{vh} &= \text{vh}\{\underline{v}\underline{v}^T\} - \text{vh}\{B^T Q(\underline{\bar{\gamma}})B\}, \\ A_{vh} &= \left[\text{vh}\{B^T \frac{\partial Q(\underline{\gamma})}{\partial \gamma_1}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}} B\}, \ldots, \text{vh}\{B^T \frac{\partial Q(\underline{\gamma})}{\partial \gamma_q}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}} B\}\right], \\ \Delta\underline{\gamma} &= \underline{\gamma} - \underline{\bar{\gamma}}, \end{aligned} \tag{4.70}$$

with $\underline{\bar{\gamma}}$ as the prior estimate of the vector of variance components. The least-squares estimate of the variance components then reads

$$\begin{aligned} \underline{\hat{\gamma}} &= \underline{\bar{\gamma}} + S^{-1}(\underline{u} - 2A_{vh}^T Q_{vh}^{-1}\text{vh}\{B^T Q(\underline{\bar{\gamma}})B\}), \\ &= \underline{\bar{\gamma}} + S^{-1}(\underline{u} - \underline{t}), \end{aligned} \tag{4.71}$$

with

$$\begin{aligned} S_{ij} &= \text{tr}\left(R\frac{\partial Q(\underline{\gamma})}{\partial \gamma_i}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}} R\frac{\partial Q(\underline{\gamma})}{\partial \gamma_j}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}}\right) \\ u_i &= \underline{y}^T R\frac{\partial Q(\underline{\gamma})}{\partial \gamma_i}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}} R\underline{y} \\ t_i &= \text{tr}\left(R\frac{\partial Q(\underline{\gamma})}{\partial \gamma_i}\Big|_{\underline{\gamma}=\underline{\bar{\gamma}}}\right), \end{aligned} \tag{4.72}$$

where use is made of Eqs. (4.44) and (B.11). Note that the redundancy matrix R is also dependent on the prior estimate $\underline{\bar{\gamma}}$. Although Eq. (4.71) equals Eq. (4.62), we cannot simplify Eq. (4.71) to Eq. (4.64), as $S\underline{\bar{\gamma}} = \underline{t}$ only holds in the linear case.

Outlier detection and robust estimation

The subject of the combination of an outlier detection on $\underline{y}$ and the estimation of the variance components will be the subject of chapter 6. It will not be useful to test for outliers in the vector $\underline{y}_{vh} = \text{vh}\{\underline{t}\underline{t}^T\}$ as outliers in the vector $\underline{y}$ will contaminate multiple elements of this vector.

4.4 Iterative Restricted Maximum Likelihood Estimator (IREML)

A drawback of the MINQUE is the computational cost of computing the elements of the S-matrix. To overcome this problem, we make use of Eqs. (4.58) and (4.64):

$$\begin{aligned} S\underline{\bar{\gamma}} &= \underline{t} \\ S\underline{\hat{\gamma}} &= \underline{u}. \end{aligned}$$

The matrix S is replaced by the diagonal matrix D in such a way that

$$D\underline{\bar{\gamma}} = \underline{t} \tag{4.73}$$

still holds [Koch, 1986]. Hence, the diagonal elements of this matrix read

$$d_i = \frac{t_i}{\bar{\gamma}_i}. \tag{4.74}$$

This diagonal matrix is then used to replace the matrix S in the computations of the BIQUE. The Iterative Restricted Maximum Likelihood Estimator (IREML) then reads

$$\hat{\gamma}_i = \frac{u_i}{t_i}\bar{\gamma}_i = \frac{\underline{y}^T R Q_i R \underline{y}}{\mathrm{tr}(RQ_i)}\bar{\gamma}_i. \tag{4.75}$$

This new estimator converges to the REML, as do the REML-fs equations. The estimator was introduced in Horn et al. [1975], although under the assumption of uncorrelated observations. The general case, in which several variance components can be estimated within one observation group, was derived by Förstner [1979]. The IREML is, therefore, also known as Förstner's estimator. As the iterations are somewhat biased, contrary to the unbiased REML-fs iterations, and they converge to unbiasedness in the limit, the estimator is often referred to as the *Almost Unbiased Estimator*. In Egeltoft [1992], an approximation of the variance of IREML is given, in which again the S-matrix is replaced by the diagonal matrix D, cf. Eq. (4.33):

$$\sigma^2_{\hat{\gamma}_i} = \frac{2}{d_i} = \frac{2\hat{\gamma}_i}{\mathrm{tr}(RQ_i)}. \tag{4.76}$$

As the elements of the S matrix does not have to be computed, the IREML is much easier to compute than the REML-fs. Another advantage of the IREML is the non-negativeness of the estimated variance components if the cofactor matrices Q_i are positive semi-definite [Rao and Kleffe, 1988].

One of the disadvantages is that the IREML, if not iterated, is generally not unbiased. Lucas [1985] found that large variance components are typically underestimated and small variance components are overestimated in the first iterations. Due to this biasedness, more iterations are needed compared to the REML-fs. Nevertheless, the computation of the IREML is still more efficient [Lucas, 1985].

In Crocetto et al. [2000] a different approach is introduced to derive the IREML. They introduce *variance factors* or *calibration factors* $\underline{v}_i$, which 'update' the variance components:

$$\hat{\gamma}_i = \hat{v}_i \cdot \bar{\gamma}_i. \tag{4.77}$$

A variance factor can be estimated by solving the system of equations:

$$L\underline{\hat{v}} = \underline{g} \tag{4.78}$$

with

$$\begin{aligned} L_{ij} &= \bar{\gamma}_i \bar{\gamma}_j S_{ij} \\ g_j &= \bar{\gamma}_j \underline{u}_j. \end{aligned}$$

The iteration process stops when $\underline{v}$ converges to the vector $[1, ..., 1]^T$. The IREML is computed by replacing the matrix L by a diagonal matrix D, in which the diagonal element d_i is the sum of the elements of the ith row of L:

$$d_i = \sum_{j=1}^{q} L_{ij} = \sum_{j=1}^{q} \bar{\gamma}_i \bar{\gamma}_j S_{ij} = \sum_{j=1}^{q} \bar{\gamma}_i \bar{\gamma}_j \text{tr}(RQ_i RQ_j) = \bar{\gamma}_i \text{tr}(RQ_i). \tag{4.79}$$

Using Eq. (4.78), we obtain

$$\hat{v}_i = \frac{\bar{\gamma}_i u_i}{\bar{\gamma}_i \text{tr}(RQ_i)} = \frac{\underline{\underline{y}}^T RQ_i R \underline{\underline{y}}}{\text{tr}(RQ_i)}. \tag{4.80}$$

This approach will be used in section 4.5 for the derivation of the IREML from the MINQUE in case of uncorrelated observations.

4.5 Uncorrelated observation groups

The computation of a global gravity model often relies on several sources of data, such as satellite tracking data, airborne gravity data and terrestrial data. One can assume that the noise of these data sets are uncorrelated to each other. Moreover, it is often legitimate to split the data of a certain type into several uncorrelated data sets. The two types of functional models, when considering uncorrelated observation groups, have already been discussed in section 3.2. In this section, we will first consider the general stochastic model for uncorrelated observations, in which the variance-covariance matrix of a certain observation group can be written as a linear combination of several cofactor matrices. Secondly, we will look at the commonly used stochastic model, in which we need to estimate a single variance component for each observation group. The equations for the MINQUE and the IREML will be derived for these different functional and stochastic models.

4.5.1 Groups with multiple variance components

When several variance components have to be estimated within the stochastic model of one observation group, the variance-covariance matrix of this group, W_i^{-1} should be written as a linear combination of the variance components and the cofactor matrices:

$$W_i^{-1} = \sum_{k=1}^{q_i} \sigma_{i,k}^2 C_{ik}. \tag{4.81}$$

Hence,

$$Q_y = \begin{bmatrix} \sum_{k=1}^{q_1} \sigma_{1,k}^2 C_{1k} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_{k=1}^{q_u} \sigma_{u,k}^2 C_{uk} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sum_{k=1}^{q_p} \sigma_{p,k}^2 C_{pk} \end{bmatrix} \tag{4.82}$$

where

$$\begin{array}{ll} q_u & = \text{Number of variance components in observation group } u \\ p & = \text{Number of observation groups} \\ \sigma_{u,k}^2 & = k\text{th variance component in observation group } u \\ C_{uk} & = m_u \times m_u \text{ matrix.} \end{array}$$

MINQUE

We will now derive the equations for the MINQUE for type I and II functional models, assuming the stochastic model of Eq. (4.82) is valid. The VC-matrix Q_y can be re-written as

$$Q_y = \sum_{u=1}^{p} \sum_{k=1}^{q_u} \sigma_{u,k}^2 Q_{uk}, \tag{4.83}$$

where

$$Q_{uk} = J_u C_{uk} J_u^T. \tag{4.84}$$

The $m \times m_u$ matrix J_u is defined as

$$J_u := \begin{bmatrix} O_{m_1} \\ \vdots \\ O_{m_{u-1}} \\ I_{m_u} \\ O_{m_{u+1}} \\ \vdots \\ O_{m_p} \end{bmatrix}, \tag{4.85}$$

where I_{m_u} is a $m_u \times m_u$ identity matrix and O_{m_i} is a $m_i \times m_i$ zero matrix. The inverse of the variance-covariance matrix can now be written as:

$$Q_y^{-1} = \sum_{r=1}^{p} J_r W_r J_r^T. \tag{4.86}$$

The system of equations to compute the MINQUE is, see Eq. (4.28):

$$S\hat{\underline{\gamma}} = \underline{u}, \tag{4.87}$$

with

$$\begin{array}{ll} \underline{\gamma} & = (\sigma^2_{1,1}...\sigma^2_{1,q_1}\sigma^2_{2,1}...\sigma^2_{2,q_2}...\sigma^2_{p,1}...\sigma^2_{p,q_p})^T \\ S_{ij} & = \text{tr}((P_A^\perp)^T Q_y^{-1} Q_{uk} (P_A^\perp)^T Q_y^{-1} Q_{vl}) \\ u_j & = \underline{y}^T (P_A^\perp)^T Q_y^{-1} Q_{vl} Q_y^{-1} P_A^\perp \underline{y}. \end{array}$$

Note that $\sigma^2_{u,k}$ is the ith element of $\underline{\gamma}$ and $\sigma^2_{v,l}$ the jth element. Using Eq. (4.84) yields

$$\begin{array}{ll} S_{ij} & = \text{tr}((P_A^\perp)^T Q_y^{-1} J_u C_{uk} J_u^T (P_A^\perp)^T Q_y^{-1} J_v C_{vl} J_v^T) \\ u_j & = \underline{y}^T (P_A^\perp)^T Q_y^{-1} J_v C_{vl} J_v^T Q_y^{-1} P_A^\perp \underline{y}. \end{array} \tag{4.88}$$

Inserting Eq. (4.86) gives:

$$\begin{array}{ll} S_{ij} & = \text{tr}((P_A^\perp)^T [\sum_{r=1}^{p} J_r W_r J_r^T] J_u C_{uk} J_u^T (P_A^\perp)^T [\sum_{s=1}^{p} J_s W_s J_s^T] J_v C_{vl} J_v^T) \\ u_j & = \underline{y}^T (P_A^\perp)^T [\sum_{r=1}^{p} J_r W_r J_r^T] J_v C_{vl} J_v^T [\sum_{s=1}^{p} J_s W_s J_s^T] P_A^\perp \underline{y}. \end{array} \tag{4.89}$$

Using the properties of the J_i matrix,

$$\begin{array}{ll} J_i^T J_j = O & \text{for all } i \neq j \\ J_i^T J_i = I_{m_i} & \end{array} \tag{4.90}$$

we can reduce Eq. (4.89) to:

$$\begin{array}{ll} S_{ij} & = \text{tr}((P_A^\perp)^T J_u W_u C_{uk} J_u^T (P_A^\perp)^T J_v W_v C_{vl} J_v^T) \\ u_j & = \underline{y}^T (P_A^\perp)^T J_v W_v C_{vl} W_v J_v^T P_A^\perp \underline{y}. \end{array} \tag{4.91}$$

When using the properties

$$\begin{array}{ll} J_u^T (P_A^\perp)^T J_v & = J_u^T J_v - W_u A_u N^{-1} A_v^T \\ J_v^T P_A^\perp \underline{y} & = \hat{\underline{e}}_v \end{array} \tag{4.92}$$

Eq. (4.91 can be re-written as

$$\begin{array}{ll} S_{ij} & = \text{tr}([J_v^T J_u - W_v A_v N^{-1} A_u^T] W_u C_{uk} [J_u^T J_v - W_u A_u N^{-1} A_v^T] W_v C_{vl}) \\ u_j & = \hat{\underline{e}}_v^T W_v C_{vl} W_v \hat{\underline{e}}_v. \end{array} \tag{4.93}$$

Hence, the equations for MINQUE when we assume a stochastic model with uncorrelated observation groups and multiple variance components to be estimated for each observation group (see Eq. (4.82)) and a type I functional model, read

$$E\{\underline{u}\} = S\underline{\gamma} \tag{4.94}$$

with

$$\begin{aligned} S_{ij}\,|_{u=v} &= \text{tr}(W_v C_{vk} W_v C_{vl}) - 2\text{tr}(N^{-1} A_v^T W_v C_{vk} W_v C_{vl} W_v A_v) + \\ &\quad + \text{tr}(N^{-1} A_v^T W_v C_{vk} W_v A_v N^{-1} A_v^T W_v C_{vl} W_v A_v) \\ S_{ij}\,|_{u\neq v} &= \text{tr}(N^{-1} A_u^T W_u C_{uk} W_u A_u N^{-1} A_v^T W_v C_{vl} W_v A_v) \\ u_j &= \underline{\hat{e}}_v^T W_v C_{vl} W_v \underline{\hat{e}}_v. \end{aligned}$$

These equations were first published in Van Loon and Kusche [2007]. We will now continue from Eq. (4.94) to derive the equations for MINQUE assuming the same stochastic model, but now with the functional model of type II. Reduction of the normal matrix N to $\bar{N}_{ss}$ results in the following equations to compute S and $\underline{u}$:

$$\begin{aligned} S_{ij}\,|_{u=v} &= \text{tr}(W_v C_{vk} W_v P^{\perp}_{A_{vv}} C_{vl}) - 2\text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vk} W_v P^{\perp}_{A_{vv}} C_{vl} W_v A_{vs}) + \\ &\quad + \text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vk} W_v P^{\perp}_{A_{vv}} A_{vs} \bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vl} W_v P^{\perp}_{A_{vv}} A_{vs}) \\ S_{ij}\,|_{u\neq v} &= \text{tr}(\bar{N}_{ss}^{-1} A_{us}^T W_u C_{uk} W_u P^{\perp}_{A_{uu}} A_{us} \bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vl} W_v P^{\perp}_{A_{vv}} A_{vs}) \\ u_j &= \underline{\hat{e}}_v^T W_v C_{vl} W_v \underline{\hat{e}}_v, \end{aligned} \tag{4.95}$$

with the projection matrices $P_{A_{ii}}$ and $P^{\perp}_{A_{ii}}$ defined as

$$\begin{aligned} P_{A_{ii}} &:= A_{ii}(A_{ii}^T W_i A_{ii})^{-1} A_{ii}^T W_i \\ P^{\perp}_{A_{ii}} &:= I - P_{A_{ii}}, \end{aligned} \tag{4.96}$$

cf. Eq. (3.8).

IREML

We will now derive the equations of the IREML using the technique suggested in section 4.4 and in Crocetto et al. [2000]. First, we will consider the functional model of type I. The variance factors v_i can easily be derived using Eq. (4.78):

$$L\underline{\hat{v}} = \underline{g}, \tag{4.97}$$

with

$$\begin{aligned} L_{ij}\,|_{u=v} &= \text{tr}(W_v \bar{\sigma}^2_{v,k} C_{vk} W_v \bar{\sigma}^2_{v,l} C_{vl}) - 2\text{tr}(N^{-1} A_v^T W_v \bar{\sigma}^2_{v,k} C_{vk} W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) + \\ &\quad + \text{tr}(N^{-1} A_v^T W_v \bar{\sigma}^2_{v,k} C_{vk} W_v A_v N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) \\ L_{ij}\,|_{u\neq v} &= \text{tr}(N^{-1} A_u^T W_u \bar{\sigma}^2_{u,k} C_{uk} W_u A_u N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) \\ g_j &= \underline{\hat{e}}_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v \underline{\hat{e}}_v. \end{aligned} \tag{4.98}$$

The diagonal elements d_j of the matrix D then read:

$$\begin{aligned} d_j &= \text{tr}(W_v[\sum_{k=1}^{q_v} \bar{\sigma}^2_{v,k} C_{vk}] W_v \bar{\sigma}^2_{v,l} C_{vl}) - \\ &\quad - 2\text{tr}(N^{-1} A_v^T W_v [\sum_{k=1}^{q_v} \bar{\sigma}^2_{v,k} C_{vk}] W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) + \\ &\quad + \text{tr}(N^{-1} [\sum_{u=1}^{p} A_u^T W_u [\sum_{k=1}^{q_u} \bar{\sigma}^2_{u,k} C_{uk}] W_u A_u] N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v). \end{aligned} \tag{4.99}$$

Noting that

$$\sum_{k=1}^{q_u} \bar{\sigma}^2_{u,k} C_{uk} = W_u^{-1}, \tag{4.100}$$

the equation reduces to

$$\begin{aligned} d_j &= \text{tr}(W_v \bar{\sigma}^2_{v,l} C_{vl}) - \\ &\quad - 2\text{tr}(N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) + \\ &\quad + \text{tr}(N^{-1} [\sum_{u=1}^{p} A_u^T W_u A_u] N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v) \\ &= \text{tr}(W_v \bar{\sigma}^2_{v,l} C_{vl}) - \text{tr}(N^{-1} A_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v A_v). \end{aligned} \tag{4.101}$$

The IREML, therefore, reads:

$$\hat{\sigma}^2_{v,l} = \frac{\underline{\hat{e}}_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v \underline{\hat{e}}_v}{\text{tr}(W_v C_{vl}) - \text{tr}(N^{-1} A_v^T W_v C_{vl} W_v A_v)}, \tag{4.102}$$

cf. [Crocetto et al., 2000]. This is the IREML when we assume a functional model of type I, i.e., without considering any local parameters. The IREML under the assumption of a type II functional model can be derived in the same way. The diagonal elements d_j can be computed by

$$\begin{aligned} d_j &= \text{tr}(P^{\perp}_{A_{vv}} \bar{\sigma}^2_{v,l} C_{vl} W_v) - \\ &\quad - 2\text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v P^{\perp}_{A_{vv}} \bar{\sigma}^2_{v,l} C_{vl} W_v A_{vs}) + \\ &\quad + \text{tr}(\bar{N}_{ss}^{-1} [\sum_{u=1}^{p} A_{us}^T W_u P^{\perp}_{A_{uu}} A_{us}] \bar{N}_{ss}^{-1} A_{vs}^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v P^{\perp}_{A_{vv}} A_{vs}). \end{aligned} \tag{4.103}$$

As

$$\bar{N}_{ss} = \sum_{u=1}^{p} A_{us}^T W_u P^{\perp}_{A_{uu}} A_{us}, \tag{4.104}$$

the IREML reads

$$\hat{\sigma}^2_{v,l} = \frac{\underline{\hat{e}}_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v \underline{\hat{e}}_v}{\text{tr}(W_v P^{\perp}_{A_{vv}} C_{vl}) - \text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v P^{\perp}_{A_{vv}} C_{vl} W_v A_{vs})}. \tag{4.105}$$

4.5.2 Groups with only one variance component

For most applications in global gravity field modelling, one only re-weights the normal matrices of each observation group due to a lack of knowledge of the error sources. The structure of the variance-covariance matrix of the observation group stays intact and a re-scaling is performed using a weighting algorithm. The variance-covariance matrix of the residuals can then be written as

$$Q_y = \begin{bmatrix} \gamma_1 C_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_j C_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \gamma_p C_p \end{bmatrix} \tag{4.106}$$

We will now re-write the VC-estimators under this assumption.

MINQUE

The equations for the MINQUE in case of a type I functional model and a stochastic model with the form of Eq. (4.106) can easily be derived using Eq. (4.94) and read

$$E\{\underline{u}\} = S\underline{\gamma} \tag{4.107}$$

with

$$\begin{aligned} S_{jj} &= \frac{1}{\bar{\gamma}_j^2}[m_j - 2\text{tr}(N^{-1}N_j) + \text{tr}(N^{-1}N_j N^{-1}N_j)] \\ S_{ij} &= \frac{1}{\bar{\gamma}_i \bar{\gamma}_j}\text{tr}(N^{-1}N_i N^{-1}N_j) \qquad \text{if } i \neq j \\ u_j &= \frac{1}{\bar{\gamma}_j^2}\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j. \end{aligned}$$

The elements of S and $\underline{u}$ under the assumption of a type II functional model read

$$\begin{aligned} S_{jj} &= \frac{1}{\bar{\gamma}_j^2}[\text{tr}(P^{\perp}_{A_{jj}}) - 2\text{tr}(\bar{N}_{ss}^{-1} A_{js}^T W_j P^{\perp}_{A_{jj}} A_{js}) + \\ &\quad + \text{tr}(\bar{N}_{ss}^{-1} A_{js}^T W_j P^{\perp}_{A_{jj}} A_{js} \bar{N}_{ss}^{-1} A_{js}^T W_j P^{\perp}_{A_{jj}} A_{js})] \\ S_{ij} &= \frac{1}{\bar{\gamma}_i \bar{\gamma}_j}\text{tr}(\bar{N}_{ss}^{-1} A_{is}^T W_i P^{\perp}_{A_{ii}} A_{is} \bar{N}_{ss}^{-1} A_{js}^T W_j P^{\perp}_{A_{jj}} A_{js}) \qquad \text{if } i \neq j \\ u_j &= \frac{1}{\bar{\gamma}_j^2}\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j. \end{aligned} \tag{4.108}$$

When we use the matrix Υ_i, see Eq. (3.25), introduced by Lucas and Dillinger [1998], we can further simplify this to

$$\begin{aligned} S_{jj} &= \frac{1}{\bar{\gamma}_j^2}[m_j - n_j - 2\text{tr}(\bar{N}_{ss}^{-1}\Upsilon_j) + \text{tr}(\bar{N}_{ss}^{-1}\Upsilon_j\bar{N}_{ss}^{-1}\Upsilon_j)] \\ S_{ij} &= \frac{1}{\bar{\gamma}_i\bar{\gamma}_j}\text{tr}(\bar{N}_{ss}^{-1}\Upsilon_i\bar{N}_{ss}^{-1}\Upsilon_j) \qquad \text{if } i \neq j \\ u_j &= \frac{1}{\bar{\gamma}_j^2}\underline{\hat{e}}_j^T C_j^{-1}\underline{\hat{e}}_j. \end{aligned} \tag{4.109}$$

IREML

Under the assumption of a type I functional model and a stochastic model of the form of Eq. (4.106), the IREML reads

$$\hat{\gamma}_j = \frac{\underline{\hat{e}}_j^T C_j^{-1}\underline{\hat{e}}_j}{m_j - \text{tr}(N^{-1}N_j)}. \tag{4.110}$$

The matrix $N^{-1}N_j$ is called the *observation group influence matrix*. The trace of this matrix is a measure of the influence of the measurements on the least-squares solution. When this trace equals the number of unknowns, the least-squares solution of the vector of unknowns only depends on this observation group. If this trace equals zero, the observation group does not contribute to the least-squares solution and could, therefore, be removed from the observation equations without changing the solution. This form of the IREML estimator has been used frequently, including [Kusche, 2003], [Xu et al., 2006], [Van Loon, 2007], and [Mayer-Gürr et al., 2007].

The IREML of $\underline{\gamma}$ in case of a type II functional model can be written as

$$\hat{\gamma}_j = \frac{\underline{\hat{e}}_j^T C_j^{-1}\underline{\hat{e}}_j}{m_i - n_i - \text{tr}(\bar{N}_{ss}^{-1}A_{js}^T W_j P_{A_{jj}}^{\perp} A_{js})} = \frac{\underline{\hat{e}}_j^T C_j^{-1}\underline{\hat{e}}_j}{m_i - n_i - \text{tr}(\bar{N}_{ss}^{-1}\Upsilon_j)} \tag{4.111}$$

The denominator of Eqs. (4.110) and (4.111) is called the *group redundancy number* r_j. Summation of all group redundancy number equals the total redundancy $m - n$. An alternative expression for the group redundancy number is

$$r_j = \text{tr}(\frac{\partial \underline{\hat{e}}_j}{\partial \underline{y}_j}) = \sum_{i=1}^{m_j} \frac{\partial \hat{e}_{ji}}{\partial y_{ji}} = \sum_{i=1}^{m_j} r_{ji}, \tag{4.112}$$

with r_{ji} the *observation redundancy number*. This number is a measure of the influence of a change in the observation to the corresponding residual. If the number is close to zero, any change in the observation will have a large effect on the solution vector, and consequently a small effect on the residual. If the observations within one observation group will have redundancy numbers close to one, the group redundancy number will be close to the number of observations. The group influence matrix is then close to zero and the observation group will have a small effect on the least-squares estimate.

4.6 Alternative weighting algorithms

We will now deal with some well-known weighting algorithms, which will not iterate to the same values as REML-fs or IREML, but have been used in global gravity field modelling. A robustified VCE will be treated in chapter 6.

4.6.1 Helmert's Variance Component Estimation

F.R. Helmert has introduced a Variance Component Estimator [Helmert, 1924], which assumes the stochastic model to be defined as

$$Q_y = \begin{bmatrix} \sigma_1^2 C_{11} & \sigma_{12} C_{12} & \cdots & \sigma_{1k} C_{1k} & \cdots & \sigma_{1p} C_{1p} \\ \sigma_{12} C_{12}^T & \sigma_2^2 C_{22} & \cdots & \sigma_{2k} C_{2k} & \cdots & \sigma_{2p} C_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{1k} C_{1k}^T & \sigma_{2k} C_{2k}^T & \cdots & \sigma_k^2 C_{kk} & \cdots & \sigma_{kp} C_{kp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{1p} C_{1p}^T & \sigma_{2p} C_{2p}^T & \cdots & \sigma_{kp} C_{kp}^T & \cdots & \sigma_p^2 C_{pp} \end{bmatrix}. \tag{4.113}$$

The (co)variance components $\sigma_1^2, \sigma_{12}, \ldots \sigma_p^2$ are collected in the $p(p+1)/2 \times 1$ vector $\underline{\gamma}$. If we set $\gamma_i = \sigma_{kl}$, the variance-covariance matrix can again be written as

$$Q_y = \sum_{i=1}^{q} \gamma_i Q_i, \tag{4.114}$$

with $q = p(p+1)/2$ and

$$Q_i := \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & C_{kl} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & C_{kl}^T & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}. \tag{4.115}$$

The inverse of the Q_y-matrix is divided into sub-matrices G_{kl}, defined as

$$Q_y^{-1} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1k} & \cdots & G_{1p} \\ G_{12}^T & G_{22} & \cdots & G_{2k} & \cdots & G_{2p} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ G_{1k}^T & G_{2k}^T & \cdots & G_{kk} & \cdots & G_{kp} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ G_{1p}^T & G_{2p}^T & \cdots & G_{kp}^T & \cdots & G_{pp} \end{bmatrix}. \tag{4.116}$$

The Q_y^{-1} matrix can now be written as

$$Q_y^{-1} = \sum_{i=1}^{q} F_i \tag{4.117}$$

with

$$F_i := \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & G_{kl} & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & G_{kl}^T & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}. \tag{4.118}$$

The matrices F_i can be written as

$$F_i = T_k^T Q_y^{-1} T_l + (1 - \delta_{kl}) T_l^T Q_y^{-1} T_k. \tag{4.119}$$

with

$$T_k := \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & I_k & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \tag{4.120}$$

and δ_{kl} the Kronecker symbol. The Helmert estimator of $\underline{\gamma}$ reads [Grafarend et al., 1980]

$$H\hat{\underline{\gamma}} = \underline{q}, \tag{4.121}$$

with

$$\begin{aligned} H_{ij} &= \text{tr}((P_A^\perp)^T F_i P_A^\perp Q_j) \\ q_j &= \underline{y}^T (P_A^\perp)^T F_j P_A^\perp \underline{y}. \end{aligned}$$

Inserting Eq. (4.119) gives

$$\begin{aligned} H_{ij} &= \text{tr}((P_A^\perp)^T [T_k Q_y^{-1} T_l + (1 - \delta_{kl}) T_l Q_y^{-1} T_k] P_A^\perp Q_j) \\ q_j &= \underline{y}^T (P_A^\perp)^T [T_r Q_y^{-1} T_s + (1 - \delta_{rs}) T_s Q_y^{-1} T_r] P_A^\perp \underline{y}. \end{aligned} \tag{4.122}$$

with $\gamma_j = \sigma_{rs}$. If and only if the observation groups are uncorrelated with each other (see Eq. (4.106)) these elements can be written as

$$\begin{aligned} H_{ij} &= \text{tr}((P_A^\perp)^T Q_y^{-1} Q_i Q_y^{-1} P_A^\perp Q_j) = S_{ij} \\ q_j &= \underline{y}^T (P_A^\perp)^T Q_y^{-1} Q_j Q_y^{-1} P_A^\perp \underline{y} = u_j, \end{aligned} \tag{4.123}$$

which is identical to the elements of MINQUE. Hence, in the case of uncorrelated observation groups, the Helmert VCE is MINQUE; see also Kelm [1978] and Grafarend et al. [1980]. The general case of Helmert's VCE, in which also covariances between observation groups are taken into account, is, however, not a MINQUE [Grafarend et al., 1980]. Many groups refer to Helmert VCE if they use MINQUE, assuming uncorrelated observation groups; see, e.g., Kizilsu and Sahin [2000].

4.6.2 The Bayesian estimate and the MAP estimate

Up till now we used a traditional approach to estimate the variance components, in which we looked at the conditional density function $p(\underline{y}|\underline{\gamma})$. This function describes the probability of the observation vector $\underline{y}$ under the assumption of the vector of variance components $\underline{\gamma}$. In Bayesian theory one makes different assumptions as one looks at the conditional (or *posterior*) density function $p(\underline{\gamma}|\underline{y})$, i.e. the probability of the vector $\underline{\gamma}$ given the observation vector $\underline{y}$. The relation between both density functions reads:

$$p(\underline{\gamma}|\underline{y}) = \frac{p(\underline{\gamma})p(\underline{y}|\underline{\gamma})}{p(\underline{y})} = \frac{p(\underline{\gamma})p(\underline{y}|\underline{\gamma})}{\int p(\underline{\gamma})p(\underline{y}|\underline{\gamma})\mathrm{d}\underline{\gamma}}. \tag{4.124}$$

With the *prior* density function $p(\underline{\gamma})$, one can accommodate for prior knowledge on $\underline{\gamma}$. However, this knowledge is not always available. Instead one can make use of a non-informative prior density function, which can be computed according to Jeffreys' invariance principle [Koch, 1990]:

$$p(\underline{\gamma}) \propto (\mathrm{det}\mathfrak{J}_{\underline{\gamma}})^{1/2}, \tag{4.125}$$

where $\mathfrak{J}_{\underline{\gamma}}$ is *Fisher's information matrix*, defined as

$$\mathfrak{J}_{\underline{\gamma}} := -E\left\{\frac{\partial^2 \ln p(y|\gamma)}{\partial\gamma_i\partial\gamma_j}\right\}. \tag{4.126}$$

Under the assumption that the observations have a multivariate normal distribution, the non-informative prior density function $p(\underline{\gamma})$ can be computed as

$$p(\underline{\gamma}) \propto (\mathrm{det}S)^{1/2}; \tag{4.127}$$

see, e.g., Eqs. (4.51), (4.56) and (4.61). Up to an unknown scale factor, the posterior density function can now be computed by:

$$p(\underline{\gamma}|\underline{y}) \propto p(\underline{\gamma})p(\underline{y}|\underline{\gamma}) \propto \frac{(\mathrm{det}S)^{1/2}}{(\mathrm{det}Q_y\mathrm{det}(A^TQ_y^{-1}A))^{1/2}}\exp\left(-\frac{1}{2}\underline{y}^TR\underline{y}\right); \tag{4.128}$$

see Eq. (4.49) for the expression of $p(\underline{y}|\underline{\gamma})$. Maximization of this posterior density function yields the *Maximum A Posteriori* (MAP) estimate. The Bayesian estimate $\underline{\gamma}_B$ for the variance components reads

$$\underline{\hat{\gamma}}_B := \int \underline{\gamma}\cdot p(\underline{\gamma}|\underline{y})\mathrm{d}\underline{\gamma}. \tag{4.129}$$

To compute this Bayesian estimate one needs the posterior density function, Eq. (4.124). Hence, integration over the product $p(\underline{\gamma})p(\underline{y}|\underline{\gamma})$ is necessary. This integration can not, however, be solved analytically. The same holds for the integration of Eq. (4.129). Numerical methods to solve this problem are suggested in Koch [1990]. However, following the argumentation in [Koch, 1990], such a Bayesian estimate will give approximately the same results as Eq. (4.64). When prior information on the density function $p(\underline{\gamma})$ is available, use can be made of this prior information in the Bayesian analysis. More on this subject can be found in Koch [1990], Ou [1991, 1993], Ou and Koch [1994] and Grodecki [1999].

An analytical expression for the MAP estimate and the Bayesian estimate is available for uncorrelated observation groups with only one variance component for each group to be estimated. The expression for the estimator, which converges to the MAP estimate, reads

$$\hat{\gamma}_{MAP,j} = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{r_j + 2} = \frac{r_j}{r_j + 2} \hat{\gamma}_j \tag{4.130}$$

where $\hat{\gamma}_j$ is the IREML. The analytical expression for the estimate, which converges to the Bayesian estimate reads, according to Ou and Koch [1994]:

$$\hat{\gamma}_{B,j} = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{r_j - 2} = \frac{r_j}{r_j - 2} \hat{\gamma}_j. \tag{4.131}$$

Hence, the MAP estimate is, for this particular stochastic model, always smaller than the IREML and the Bayesian estimate is always larger than the IREML. If the group redundancy number r_j is large, all three estimators are approximately equal.

4.6.3 Lerch's subset solution method

In Lerch's method [Lerch, 1989,1991], a solution using all observation groups is compared to the solution in which one observation group is omitted. The difference between both solutions should be in agreement with the expectation of the difference, based on the stochastic model. In Lerch [1989, 1991], the observations are assumed to be uncorrelated. Those equations will now be generalized under the assumption of a block-diagonal variance-covariance matrix of which only one variance component is to be estimated for each block, see Eq. (4.106). Hence, Lerch's method can not be used in the computation of multiple variance components per observation group. The least-squares solution using the complete data set, $\underline{\hat{x}}$, is given by (Eq. 3.18). Omitting observation group j results in a different least-squares solution, $\underline{\hat{x}}_{[j]}$:

$$N_{[j]} \underline{\hat{x}}_{[j]} = \underline{h}_{[j]} \quad ; \quad Q_{\hat{x}_{[j]}} = N_{[j]}^{-1}, \tag{4.132}$$

in which

$$\begin{aligned} N_{[j]} &= (\sum_{\substack{i=1 \\ i\neq j}}^{p} N_i) \\ \underline{h}_{[j]} &= (\sum_{\substack{i=1 \\ i\neq j}}^{p} \underline{h}_i). \end{aligned}$$

Hence

$$\begin{aligned} N &= N_{[j]} + N_j \\ \underline{h} &= \underline{h}_{[j]} + \underline{h}_j. \end{aligned} \tag{4.133}$$

If the influence of the data set $\underline{y}_j$ is small compared to the overall data set, we find

$$\begin{aligned} N^{-1} &\approx N_{[j]}^{-1} - N_{[j]}^{-1} N_j N_{[j]}^{-1} \\ \underline{\hat{x}} &\approx \underline{\hat{x}}_{[j]} + N_{[j]}^{-1} \underline{\tilde{h}}_j. \end{aligned} \tag{4.134}$$

Lerch's method compares the unweighted square of the difference between the subset solution and the complete solution, i.e., $(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})$, with its expected value, based on the stochastic model. This expectation can be computed by

$$E\{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})\} = \text{tr}(D\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\}) \tag{4.135}$$

if $E\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\} = \underline{0}$. The variance-covariance matrix of this difference, $D\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\}$, can be computed by

$$D\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\} = D\{\underline{\hat{x}}_{[j]}\} - 2C\{\underline{\hat{x}}_{[j]}, \underline{\hat{x}}\} + D\{\underline{\hat{x}}\}, \tag{4.136}$$

with

$$\begin{aligned} C\{\underline{\hat{x}}_{[j]}, \underline{\hat{x}}\} &= N_{[j]}^{-1} C\{\underline{h}_{[j]}, \underline{h}\} N^{-1} \\ &= N_{[j]}^{-1} C\{\underline{h}_{[j]}, \underline{h}_{[j]}\} N^{-1} \\ &= N_{[j]}^{-1} N_{[j]} N^{-1} \\ &= N^{-1}. \end{aligned} \tag{4.137}$$

This results in

$$D\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\} = N_{[j]}^{-1} - N^{-1}. \tag{4.138}$$

The variance factor can then be computed by comparing the square of the differences with its expectation:

$$\hat{v}_j = \frac{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})}{E\{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})\}} = \frac{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})}{\text{tr}(N_{[j]}^{-1} - N^{-1})}. \tag{4.139}$$

This factor is then used to update the variance components:

$$\hat{\gamma}_j = \hat{v}_j \cdot \bar{\gamma}_j. \tag{4.140}$$

A different observation group is then omitted and the variance factor is computed for this observation group. This algorithm is repeated until all variance factors approximate unity. At this point, all data sets are balanced in the sense that their actual influence on the solution is in agreement with its expectation, the latter being based on the stochastic model. Lerch's estimator is an approximately unbiased estimator if the approximations of Eq. (4.134) hold:

$$\begin{aligned} E\{\hat{\gamma}_j\} &= \bar{\gamma}_j E\{\hat{v}_j\} = \bar{\gamma}_j \frac{E\{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})\}}{\mathrm{tr}(N_{[j]}^{-1} - N^{-1})} \\ &= \bar{\gamma}_j \frac{\mathrm{tr}(D\{\underline{\hat{x}}_{[j]} - \underline{\hat{x}}\})}{\mathrm{tr}(N_{[j]}^{-1} - N^{-1})} \approx \bar{\gamma}_j \frac{\mathrm{tr}(D\{N_{[j]}^{-1}\underline{\tilde{h}}_j\})}{\mathrm{tr}(N_{[j]}^{-1} N_j N_{[j]}^{-1})} \\ &= \frac{\mathrm{tr}(N_{[j]}^{-1} D\{\underline{\tilde{h}}_j\} N_{[j]}^{-1})}{\mathrm{tr}(N_{[j]}^{-1} A_j^T C_j^{-1} A_j N_{[j]}^{-1})} = \frac{\mathrm{tr}(N_{[j]}^{-1} \gamma_j A_j^T C_j^{-1} A_j N_{[j]}^{-1})}{\mathrm{tr}(N_{[j]}^{-1} A_j^T C_j^{-1} A_j N_{[j]}^{-1})} = \gamma_j. \end{aligned} \tag{4.141}$$

This is true if the influence of the omitted observation group is small compared to the overall data set. However, small observation groups can cause instability problems [Lerch et al., 1994] when they do not add extra information. The least-squares solution and the normal matrix of the complete data set will then be almost equal to the data set omitting this observation group, which results in an unrealistic calibration. To overcome this problem, several similar small observation groups have to be combined to one larger observation group. Alternatively, one leaves their stochastic models unaltered.

The GEM-T2 [Marsh et al., 1990] and GEM-T3 [Lerch et al., 1994] global gravity models used this subset solution method to calibrate the different data groups. Three groups of data, namely satellite tracking data, surface gravity data and radar altimeter data, were calibrated against each other as a test in the GEM-T3 solution. The calibration of the surface gravity data and the altimeter data worked well. However, the results were bad when calibrating the satellite tracking data as the satellite data is strong in the low-degree part of the spectrum, compared to the surface gravity and radar altimeter data, which mostly contribute to the medium and high-frequency part [Lerch et al., 1994].

Eq. (4.139) calibrates the data sets with respect to all unknowns. However, sometimes it would be better to omit some unknowns in the calibration procedure. If for example the goal is to estimate the lower-degree potential coefficients, one should only calibrate these coefficients and remove other unknowns from the calibration process [Lemoine et al., 1998]

In Lerch et al. [1993, 1994], an alternative subset solution method is proposed, in which the actual residuals ($y_k - A_k \underline{\hat{x}}_{[k]}$) are compared with the projected stochastic model ($A_k^T N_{[k]} A_k$) together with the uncertainties in the tides, station coordinates and polar motion. The method was tested for the Starlette and Ajisai satellite data, which were used in the computation of the GEM-T3 global gravity model. It was shown that these satellite data were well-calibrated.

4.6.4 Generalized Cross-Validation (GCV)

An unstable system of normal equations can be stabilized by adding pseudo-observations to the functional and stochastic model, i.e.

$$\underline{0} = \underline{x} + \underline{e}_K \quad ; \quad E\{\underline{e}_K\} = \underline{0} \quad ; \quad D\{\underline{e}_K\} = (\alpha K)^{-1}, \tag{4.142}$$

with K a symmetric positive-definite regularization matrix. This stabilization is a form of Tikhonov regularization, see e.g., section (2.1.2). The regularized least-squares estimator then reads

$$\underline{\hat{x}}_\alpha = (A^T Q_y^{-1} A + \alpha K)^{-1} A^T Q_y^{-1} \underline{y}. \tag{4.143}$$

As the pseudo-observations can be seen as an extra data set, we can obtain an estimate of the regularization parameter using VCE, e.g., IREML as follows

$$\hat{\alpha} = \frac{n - \bar{\alpha}\text{tr}((A^T Q_y^{-1} A + \bar{\alpha} K)^{-1} K)}{\underline{\hat{x}}_\alpha^T K \underline{\hat{x}}_\alpha}. \tag{4.144}$$

The estimation of such a regularization parameter can be unstable. An alternative method is the (Generalized) Cross-Validation.

The Cross-Validation (CV) method is based on the leave-one-out principle, in which an observation value is compared with the predicted value, using all other observations. If one assumes a scaled identity matrix as the variance-covariance matrix, the CV-estimate of α is the α which minimizes this difference [Golub et al., 1979]:

$$\hat{\alpha}_{CV} = \arg\min \frac{1}{m} \sum_{i=1}^{m} \left((y_i - A\underline{\hat{x}}_\alpha^{[i]})_i \right)^2. \tag{4.145}$$

This can be re-written as

$$\hat{\alpha}_{CV} = \arg\min \frac{1}{m} \sum_{i=1}^{m} \frac{((y_i - A\underline{\hat{x}}_\alpha)_i)^2}{(1 - V_{ii})^2}, \tag{4.146}$$

with $V = A(A^T Q_y^{-1} A + \alpha K)^{-1} A^T Q_y^{-1}$. This is a much faster algorithm, as one does not need to compute m estimates $\underline{\hat{x}}_\alpha^{[i]}$. If correlations are taken into account for the observations and if the denominators are replaced by their mean value $\text{tr}(I - V)/m$, we obtain the *Generalized* Cross-Validation estimate; see, e.g., Kusche and Klees [2002]

$$\hat{\alpha}_{GCV} = \arg\min \frac{m(\underline{y} - A\underline{\hat{x}}_\alpha)^T Q_y^{-1} (\underline{y} - A\underline{\hat{x}}_\alpha)}{(m - \text{tr}((A^T Q_y^{-1} A + \bar{\alpha} K)^{-1} A^T Q_y^{-1} A))^2}. \tag{4.147}$$

4.6.5 (Iterative) Maximum Likelihood Estimator (MLE and IMLE)

The equations for MINQUE converge to the Restricted Maximum Likelihood (REML) if one assumes normally distributed data. Instead of setting the estimate of the vector of

unknowns $\underline{x}$ as a vector of *free* parameters, as was done in the REML, we will now *fix* this vector.

Again, we will assume the probability density function of Eq. (4.47). But now we take the variance-covariance matrix Q_y as a fixed estimate (true values) and only differentiate with respect to $\underline{x}$ to find maximum likelihood:

$$\frac{\partial \ln p(\underline{y}|\underline{x},\underline{\gamma})}{\partial \underline{x}} = A^T Q_y^{-1} \underline{y} - A^T Q_y^{-1} A \underline{x}. \tag{4.148}$$

Setting this equation equal to zero results in the well-known Weighted Least-Squares solution of the vector of unknowns (see, e.g., Eq. (3.6)):

$$A^T Q_y^{-1} A \underline{\hat{x}} = A^T Q_y^{-1} \underline{y}.$$

Differentiation of the function $\ln p(\underline{y}|\underline{x},\underline{\gamma})$ with respect to the variance component γ_i reads

$$\frac{\partial \ln p(\underline{y}|\underline{x},\underline{\gamma})}{\partial \gamma_i} = -\frac{1}{2}\text{tr}(Q_y^{-1}Q_i) + \frac{1}{2}(\underline{y} - A\underline{x})^T Q_y^{-1} Q_i Q_y^{-1} (\underline{y} - A\underline{x}). \tag{4.149}$$

where use is made of the properties (Eq. (B.4) and (B.5)) of a symmetric matrix; see also Koch [1986,1990]. Replacing $\underline{x}$ by its MLE $\underline{\hat{x}}$ results in the equation for the MLE of the variance component γ_i:

$$-\text{tr}(Q_y^{-1}Q_i) + \underline{y}^T R Q_i R \underline{y} = 0, \tag{4.150}$$

where R is Rao's matrix; see Eq. (4.25). The equation can be solved iteratively, e.g., with Newton-Raphson or Fisher scoring. However, there are some disadvantages to using the MLE to estimate the variance components. It takes the least-squares estimate $\underline{\hat{x}}$ as the *true* value for $\underline{x}$ and does not take into account the degrees of freedom that are involved in estimating $\underline{\hat{x}}$. Moreover, the MLE is, in general, biased and has no minimum variance [Searle et al., 1992].

If we now assume a stochastic model of Eq. (4.106), i.e. consider one variance component for each observation group, the MLE reads at convergence

$$\gamma_j = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{m_j}. \tag{4.151}$$

In Yuan [1991], this equation is used at each iteration:

$$\hat{\gamma}_j = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{m_j}. \tag{4.152}$$

We will call this estimator the *Iterative Maximum Likelihood Estimator (IMLE)*. As the estimates converge to the estimates of the Maximum Likelihood Estimator, the IMLE has the same disadvantages as the MLE. The numerical computation of the estimator is however very fast and shall be tested against the other estimators in chapter 7. The estimator was used in several global gravity field models, e.g., TEG-3 [Tapley et al., 1997] and GGM02S [Tapley et al., 2005].

4.6.6 External calibration

Many global gravity models have been derived by comparing the results to external (independent) data. One could choose the data weights by trial and error to minimize the difference between the solution and the independent data. In Lerch et al. [1993], a more sophisticated method is introduced to externally calibrate a solution. The solution is compared to SLR observations. The difference is, similar to Lerch's subset solution method, compared to the expected difference, based on the stochastic models. If necessary, the stochastic model of the global gravity solution is rescaled by the calibration factor of the external calibration.

4.7 Summary

Observations are in general corrupted by errors. With a proper modelling of the functional model, one tries to reduce other effects than the stochastic observation noise from the vector of residuals $\underline{e}$, i.e. the difference between the observation vector and its expected value. Those other effects could include systematic errors, truncation errors, errors in the background models (e.g., tidal or atmospheric models) or inconsistencies with other data sets. The stochastic properties of the remaining residual vector are defined in the stochastic model, in which the variance-covariance matrix Q_y defines the dispersion of the residual vector.

However, this dispersion is in general not known beforehand and one needs to estimate the Q_y-matrix. To do this, we choose to write this matrix as a linear combination of several (cofactor) matrices, in which the *variance components* are the linear coefficients. The division of the variance-covariance matrix in its components is however rather arbitrary and non-unique. One should base this linear combination on the properties of the different error sources contributing to the observation noise.

In a least-squares approach, one tries to minimize the residual square sum $\underline{e}^T Q_y^{-1} \underline{e}$. A proper modelling of the variance-covariance matrix is therefore needed to compute the least-squares solution of the vector of unknowns and its variance-covariance matrix.

The *Variance Components Estimation (VCE)* can be done using several estimators, each with its own properties. We have discussed some estimators, which under normality converge to the same estimate. This estimate is unbiased, translation invariant, has minimum variance, converge to maximum likelihood and is based on a weighted least-squares approach. Within this group of estimators, the Iterative Restricted Maximum Likelihood Estimator (IREML) is somewhat faster than the other estimators, like MINQUE, but it is biased in each iteration and requires more iterations till convergence.

We have derived equations to simplify the discussed estimators in case of uncorrelated observation groups with either multiple variance components or just a single variance component to be estimated within one observation group. This is common practice in global gravity field modelling, as many independent data sets (e.g., satellite data, marine data, airborne data and terrestrial data) are involved.

Alternatively, we have discussed some variance components estimators, which are well-known in the weight determination of global gravity models, e.g., Helmert VCE and Lerch's subset solution method. However, they do not fulfill in general the properties of unbiasedness, maximum likelihood and minimum variance.

5

Monte Carlo implementation

The estimation of the variance components (e.g., MINQUE, IREML) involves the computation of the trace of a matrix. This matrix is the result of the multiplication of several matrices, including the inverse of the accumulated normal matrix. As this inverse is generally not available for large-scale problems (which are usually solved by, e.g., Cholesky factorization), one cannot compute the trace directly. In this chapter, we will use the alternative of the *stochastic trace estimation* (STE), in which we do not need the inverse of the normal matrix to obtain an estimate of the trace. We will first introduce this approach and then use STE for several estimators discussed in the previous chapter, including Lerch's subset solution method. It will be shown that this leads to very efficient Monte Carlo algorithms. One can use any least-squares software package and use it as a black box to solve for the variance components. In this way, the extra computational effort involved in VCE remains limited. Although STE could also be beneficial for the computation of Helmert's VCE and GCV, we will only focus on the MINQUE, IREML, and Lerch's estimator. The use of STE in GCV is discussed in Kusche and Klees [2002].

5.1 Stochastic Trace Estimation (STE)

The STE is based on one of the properties of a square matrix, i.e. [Girard, 1989]

$$\mathrm{tr}(T) = E\{\underline{z}^T T \underline{z}\}, \tag{5.1}$$

where $\underline{z}$ is a random vector with $E\{\underline{z}\} = \underline{0}$ and $D\{\underline{z}\} = I$. In Girard [1989], it was stated that this property holds for *symmetric positive-definite* matrices. We can show

however, that this property holds for every square matrix:

$$\begin{aligned} E\{\underline{z}^T T \underline{z}\} &= E\{\text{tr}(\underline{z}^T T \underline{z})\} \\ &= E\{\text{tr}(T \underline{z}\underline{z}^T)\} \\ &= E\{[T \underline{z}\underline{z}^T]_{11} + [T \underline{z}\underline{z}^T]_{22} + \ldots\} \\ &= E\{T_{11}[\underline{z}\underline{z}^T]_{11} + T_{12}[\underline{z}\underline{z}^T]_{21} + \ldots\} \\ &= T_{11} E\{[\underline{z}\underline{z}^T]_{11}\} + T_{12} E\{[\underline{z}\underline{z}]_{21}\} + \ldots \\ &= [T E\{\underline{z}\underline{z}^T\}]_{11} + [T E\{\underline{z}\underline{z}^T\}]_{22} + \ldots \\ &= \text{tr}(T E\{\underline{z}\underline{z}^T\}) \\ &= \text{tr}(T(D\{\underline{z}\} + E\{\underline{z}\}E\{\underline{z}\})) \\ &= \text{tr}(T). \end{aligned} \tag{5.2}$$

The expectation operator $E\{.\}$ can be approximated by the average over r realizations (sample mean):

$$E\{\underline{z}^T T \underline{z}\} \approx \frac{1}{r} \sum_{k=1}^{r} \underline{z}_k^T T \underline{z}_k. \tag{5.3}$$

In several experiments (see, e.g., Girard [1989, 1995] and Golub and Von Matt [1997]) it was found that one realization of $\underline{z}$ is the best compromise between accuracy and computational costs, when dealing with large-scale problems. Therefore only one realization will be used in the following derivations. It is, however, possible that multiple realizations are necessary in some exceptional cases.

Up till now, no assumption has been made on the probability density function of the vector $\underline{z}$. When using the multivariate normal distribution, the variance of the trace estimate can be bounded by Girard [1989]

$$V(\underline{z}^T T \underline{z}) = 2\text{tr}(T^2). \tag{5.4}$$

The variance of the trace estimate attains its minimum if a multivariate discrete binary probability function is chosen as the probability function for $\underline{z}$. Then each sample of the vector $\underline{z}$ takes the value -1 or 1 with equal probability. The variance is now bounded by Hutchinson [1990]

$$V(\underline{z}^T T \underline{z}) = 2 \sum_{k \neq l}^{n} (T_{kl})^2 < 2\text{tr}(T^2) - \frac{2}{n}(\text{tr}(T))^2 < 2\text{tr}(T^2). \tag{5.5}$$

As we will show in the next sections, STE can be used as a Monte Carlo method to estimate the trace of a matrix. The equations will be derived in such a way that the computation of the variance components does not require a new Cholesky factorization or inversion of a large normal matrix. One can simply derive randomized pseudo-observations and insert them into existing software, which is then considered as a 'black box'. The output will then be used to compute the variance components.

Koch and Kusche [2002] used for the first time Monte Carlo VCE (MC-VCE) to compute the regularization parameter. Kusche [2003] provides us with more information and a test with a simulated GOCE orbit. We will derive the Monte Carlo variants of the MINQUE, IREML and Lerch estimators and show the equations for the functional models of type I and type II.

5.2 Monte Carlo MINQUE (MC-MINQUE)

We will now derive algorithms to compute the MINQUE using STE. The general form of the MINQUE, i.e., Eq. (4.28),

$$E\{\underline{u}\} = S\underline{\gamma}$$

with

$$\begin{aligned} S_{ij} &= \text{tr}(RQ_iRQ_j) \\ u_j &= \underline{y}^T RQ_jR\underline{y} \end{aligned}$$

will not be implemented as this form is not applicable for large-scale problems due to the inversion of the variance-covariance matrix Q_y in the computation of R. We will therefore focus on the situation of uncorrelated observation groups.

5.2.1 Groups with multiple variance components

Functional model of type I

In the situation of uncorrelated observation groups, in which the stochastic model of an individual observation group can be written as a linear combination of several cofactor matrices, the elements of S and $\underline{u}$ can be written as (Eq. (4.94)):

$$\begin{aligned} S_{ij}\,|_{u=v} &= \text{tr}(W_vC_{vk}W_vC_{vl}) - 2\text{tr}(N^{-1}A_v^TW_vC_{vk}W_vC_{vl}W_vA_v) + \\ &\quad + \text{tr}(N^{-1}A_v^TW_vC_{vk}W_vA_vN^{-1}A_v^TW_vC_{vl}W_vA_v) \\ S_{ij}\,|_{u\neq v} &= \text{tr}(N^{-1}A_u^TW_uC_{uk}W_uA_uN^{-1}A_v^TW_vC_{vl}W_vA_v) \\ u_j &= \underline{\hat{e}}_v^TW_vC_{vl}W_v\underline{\hat{e}}_v. \end{aligned}$$

The different traces can be estimated by STE. The first trace $\text{tr}(W_vC_{vk}W_vC_{vl})$ can be computed explicitly as the inversion of the normal matrix is not involved. The computation of the matrix W_v needs a matrix inversion of size $m_v \times m_v$. However, this matrix is already computed in the least-squares estimation. The second trace can be estimated by

$$\begin{aligned} \text{tr}(N^{-1}A_v^TW_vC_{vk}W_vC_{vl}W_vA_v) &= \text{tr}(N^{-1}A_v^TW_vC_{vk}D_vD_v^TC_{vl}W_vA_v) \\ &= \text{tr}(D_v^TC_{vl}W_vA_vN^{-1}A_v^TW_vC_{vk}D_v) \\ &\approx \underline{z}_v^TD_v^TC_{vl}W_vA_vN^{-1}A_v^TW_vC_{vk}D_v\underline{z}_v \\ &= \underline{z}_v^TD_v^TC_{vl}W_vA_v\underline{\hat{q}}_{vk}, \end{aligned} \tag{5.6}$$

where use is made of the decomposition (e.g., Cholesky)

$$W_v = D_vD_v^T, \tag{5.7}$$

and the solution of the randomized normal equations

$$N\underline{\hat{q}}_{vk} = A_v^TW_vC_{vk}D_v\underline{z}_v. \tag{5.8}$$

If one replaces the vector of observations as

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_{v-1} \\ \underline{y}_v \\ \underline{y}_{v+1} \\ \vdots \\ \underline{y}_p \end{bmatrix} \rightarrow \underline{y}^q_{vk} = \begin{bmatrix} \underline{0} \\ \vdots \\ \underline{0} \\ C_{vk} D_v \underline{z}_v \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}, \tag{5.9}$$

one can use existing software as a black box to compute the vector $\hat{\underline{q}}_{vk}$ and consequently the variance components. Furthermore,

$$\text{tr}(N^{-1} A_v^T W_v C_{vk} W_v A_v N^{-1} A_v^T W_v C_{vl} W_v A_v) \approx \hat{\underline{q}}_{vl}^T A_v^T W_v C_{vk} W_v A_v \hat{\underline{q}}_{vl} \tag{5.10}$$

and

$$\text{tr}(N^{-1} A_u^T W_u C_{uk} W_u A_u N^{-1} A_v^T W_v C_{vl} W_v A_v) \approx \hat{\underline{q}}_{vl}^T A_u W_u C_{uk} W_u A_u \hat{\underline{q}}_{vl}, \tag{5.11}$$

The algorithm can be found in figure 5.1.

Hence, one needs to factorize W_v to compute the pseudo-observations. This vector will then be inserted into any least-squares software to compute the vectors $\hat{\underline{q}}_{uk}$ instead of the vector of unknowns $\hat{\underline{x}}$. After the least-squares software, one only needs to do some matrix-vector multiplications to get the required variance components. These new estimates are then used to re-compute the VC-matrix, which will be inserted in the least-squares software. The Monte Carlo variant of MINQUE was for the first time derived in Van Loon and Kusche [2007]. The proposed algorithm in this thesis is a slight modification of the algorithm proposed in that paper.

Functional model of type II

A functional model of type II takes local parameters into account. The equations to derive the MINQUE in case of uncorrelated observation groups with multiple variance components for each group and a functional model of type II read (Eq. (4.95)):

$$\begin{aligned} S_{ij}\,|_{u=v} &= \text{tr}(C_{vk} W_v P^{\perp}_{A_{vv}} C_{vl} W_v) - 2\text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vk} W_v P^{\perp}_{A_{vv}} C_{vl} W_v A_{vs}) + \\ &\quad + \text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vk} W_v P^{\perp}_{A_{vv}} A_{vs} \bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vl} W_v P^{\perp}_{A_{vv}} A_{vs}) \\ S_{ij}\,|_{u\neq v} &= \text{tr}(\bar{N}_{ss}^{-1} A_{us}^T W_u C_{uk} W_u P^{\perp}_{A_{uu}} A_{us} \bar{N}_{ss}^{-1} A_{vs}^T W_v C_{vl} W_v P^{\perp}_{A_{vv}} A_{vs}) \\ u_j &= \hat{\underline{e}}_v^T W_v C_{vl} W_v \hat{\underline{e}}_v, \end{aligned}$$

One could now start with replacing the traces by stochastic trace estimators. It would however be easier to take the algorithm of figure 5.1 and to make some slight modifications. The output of the black box algorithm will now be used assuming a different functional model. We therefore have to replace A_v by $[\ 0\ \ \ldots\ \ 0\ \ A_{vv}\ \ 0\ \ \ldots\ \ 0\ \ A_{vs}\ ]$ in the computation of the stochastic models. Note that we need all elements of the vectors $\hat{\underline{q}}_{uk}$.

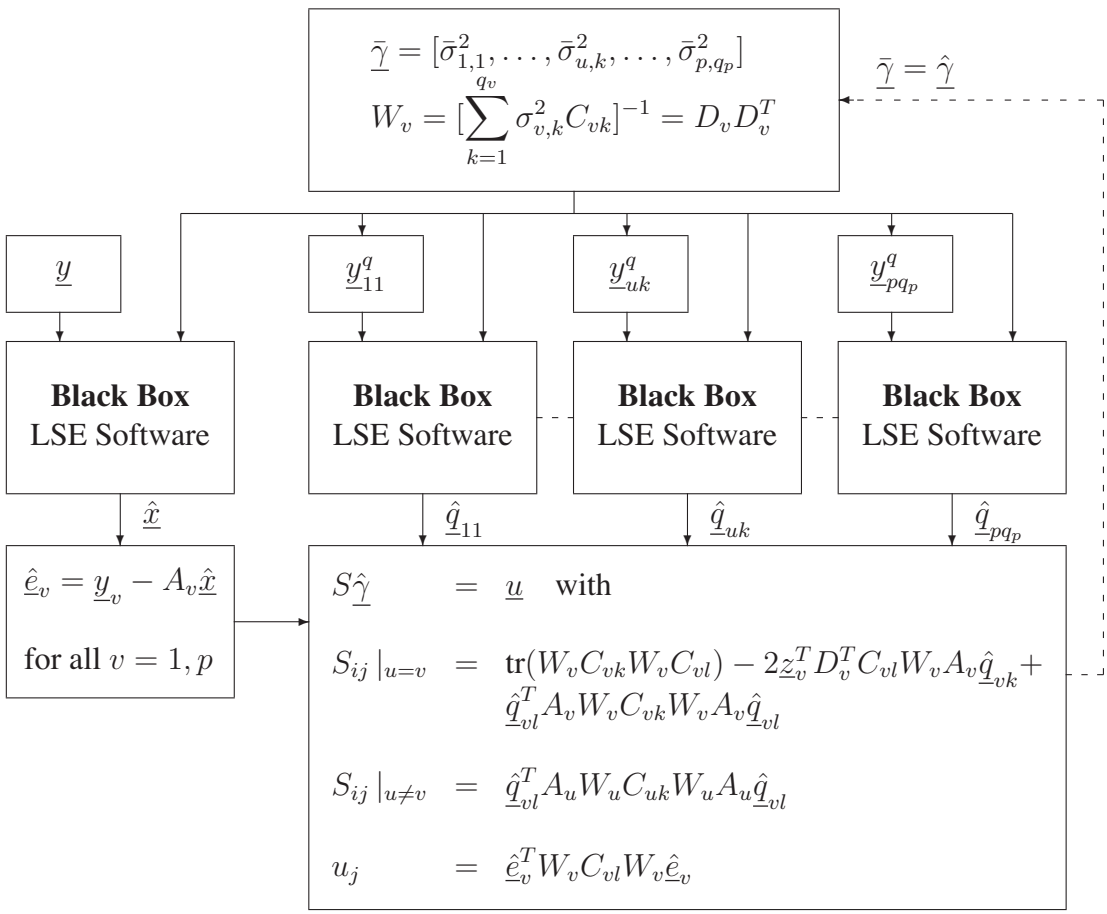


Fig. 5.1: The algorithm to estimate the vector of unknowns $\underline{\hat{x}}$ and the variance components $\underline{\hat{\gamma}}$ using MINQUE with uncorrelated observation groups, multiple variance components per group, and a functional model of type I.

5.2.2 Groups with only one variance component

Functional model of type I

If the variance-covariance matrices of the individual observation groups are known up to a certain scaling parameter, one only needs to estimate one variance component per observation group. The elements of MINQUE then read (see Eq. (4.107)):

$$\begin{aligned} S_{jj} &= \frac{1}{\bar{\gamma}_j^2}[m_j - 2\text{tr}(N^{-1}N_j) + \text{tr}(N^{-1}N_jN^{-1}N_j)] \\ S_{ij} &= \frac{1}{\bar{\gamma}_i\bar{\gamma}_j}\text{tr}(N^{-1}N_iN^{-1}N_j) \qquad \text{if } i \neq j \\ u_j &= \frac{1}{\bar{\gamma}_j^2}\hat{\underline{e}}_j^T C_j^{-1}\hat{\underline{e}}_j. \end{aligned}$$

We can simply derive the equations for the STE of the traces, i.e., for the trace of the group influence matrix $N^{-1}N_j$:

$$\begin{aligned} \text{tr}(N^{-1}N_j) &= \text{tr}(N^{-1}A_j^T W_j \bar{\gamma}_j C_j W_j A_j) \\ &\approx \bar{\gamma}_j \underline{z}_j^T G_j^T W_j A_j N^{-1} A_j^T W_j G_j \underline{z}_j \\ &= \bar{\gamma}_j \underline{z}_j^T G_j^T W_j A_j \hat{\underline{q}}_j \\ &= \underline{z}_j^T G_j^T C_j^{-1} A_j \hat{\underline{q}}_j \end{aligned} \tag{5.12}$$

with $C_j = G_j G_j^T$ and

$$N\hat{\underline{q}}_j = A_j^T W_j G_j^T \underline{z}_j = A^T W \underline{y}_j^q \tag{5.13}$$

in which $\underline{y}_j^q$ is defined by

$$\underline{y}_j^q = \begin{bmatrix} \underline{0} \\ \vdots \\ \underline{0} \\ G_j\underline{z}_j \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}. \tag{5.14}$$

Note that the variance-covariance matrix of $\underline{y}_j^q$ actually reads (propagation law of variances):

$$Q_y = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & C_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \tag{5.15}$$

which is the cofactor matrix corresponding to γ_j. However, we will assume

$$Q_y = \begin{bmatrix} \gamma_1 C_1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \gamma_j C_j & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \gamma_p C_p \end{bmatrix} \tag{5.16}$$

as its variance-covariance matrix in the black-box algorithm. The STE of the other trace $\text{tr}(N^{-1}N_iN^{-1}N_j)$ can be derived in a similar way

$$\text{tr}(N^{-1}N_iN^{-1}N_j) \approx \bar{\gamma}_j \hat{\underline{q}}_j^T A_i^T W_i A_i \hat{\underline{q}}_j = \frac{\bar{\gamma}_j}{\bar{\gamma}_i} \hat{\underline{q}}_j^T A_i^T C_i^{-1} A_i \hat{\underline{q}}_j \tag{5.17}$$

This algorithm can be found in figure 5.2.

Note that in this algorithm the factorization step (this time: $C_i = G_j G_j^T$) is only needed once, as C_j does not change in the iteration process. Moreover, one does not need to change the randomized input vector. This vector $\underline{y}_j^q$ also stays intact.

Functional model of type II

If the functional model is of type II, the elements of MINQUE can be estimated by Eq. (4.108):

$$\begin{aligned} S_{jj} &= \frac{1}{\bar{\gamma}_j^2}[\text{tr}(P_{A_{jj}}^{\perp}) - 2\text{tr}(\bar{N}_{ss}^{-1} A_{js}^T W_j P_{A_{jj}}^{\perp} A_{js}) + \\ &\quad + \text{tr}(\bar{N}_{ss}^{-1} A_{js}^T W_j P_{A_{jj}}^{\perp} A_{js} \bar{N}_{ss}^{-1} A_{js}^T W_j P_{A_{jj}}^{\perp} A_{js})] \\ S_{ij} &= \frac{1}{\bar{\gamma}_i \bar{\gamma}_j} \text{tr}(\bar{N}_{ss}^{-1} A_{is}^T W_i P_{A_{ii}}^{\perp} A_{is} \bar{N}_{ss}^{-1} A_{js}^T W_j P_{A_{jj}}^{\perp} A_{js}) \qquad \text{if } i \neq j \\ u_j &= \frac{1}{\bar{\gamma}_j^2} \hat{\underline{e}}_j^T C_j^{-1} \hat{\underline{e}}_j. \end{aligned}$$

Instead of replacing these traces using STE, we can use the algorithm of figure 5.2 and slightly modify it, as was also done in the case of multiple variance components per observation group. One simply replaces A_i with $[\ 0 \ \ldots \ 0 \ A_{ii} \ 0 \ \ldots \ 0 \ A_{is}\ ]$. If the reduced normal matrices Υ_j are available one can further simplify the equations. Use is made of the system of normal equations to compute the vector $\hat{\underline{q}}_i$:

$$\begin{bmatrix} N_{11} & \cdots & 0 & \cdots & 0 & N_{1s} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & N_{ii} & \cdots & 0 & N_{is} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & N_{pp} & N_{ps} \\ N_{s1} & \cdots & N_{si} & \cdots & N_{sp} & N_{ss} \end{bmatrix} \begin{bmatrix} \hat{\underline{q}}_{i1} \\ \vdots \\ \hat{\underline{q}}_{ii} \\ \vdots \\ \hat{\underline{q}}_{ip} \\ \hat{\underline{q}}_{is} \end{bmatrix} = \begin{bmatrix} \underline{0} \\ \vdots \\ A_{ii}^T W_i G_i \underline{z}_i \\ \vdots \\ \underline{0} \\ A_{is}^T W_i G_i \underline{z}_i \end{bmatrix}. \tag{5.18}$$

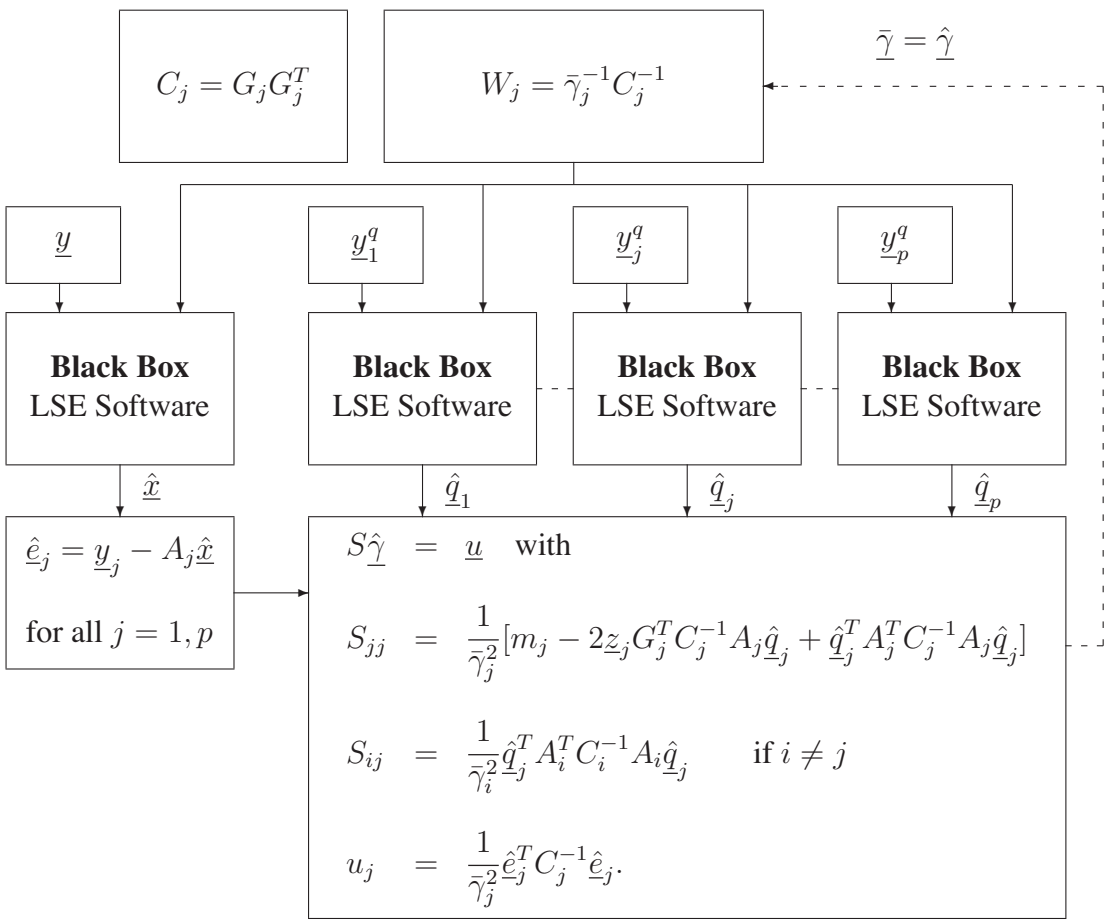


Fig. 5.2: The algorithm to estimate the vector of unknowns $\underline{\hat{x}}$ and the variance components $\underline{\hat{\gamma}}$ using MINQUE with uncorrelated observation groups, one variance component per group, and a functional model of type I.

Hence, the global vector $\underline{\hat{q}}_{is}$ is computed by solving

$$\bar{N}_{ss}\underline{\hat{q}}_{is} = A_{is}^T W_i G_i \underline{z}_i - N_{is}^T N_{ii}^{-1} A_{ii}^T W_i G_i \underline{z}_i. \tag{5.19}$$

The local vectors are obtained by solving

$$\begin{array}{lll} N_{ii}\underline{\hat{q}}_{ii} & = -N_{is}\underline{\hat{q}}_{is} + A_{ii}^T W_i G_i \underline{z}_i & \\ N_{jj}\underline{\hat{q}}_{ij} & = -N_{js}\underline{\hat{q}}_{is} & \text{if } i \neq j. \end{array} \tag{5.20}$$

Both traces can now be estimated by

$$\begin{array}{ll} \text{tr}(N^{-1}N_i N^{-1}N_j) & \approx \bar{\gamma}_i \cdot (A_{jj}\underline{\hat{q}}_{ij} + A_{js}\underline{\hat{q}}_{is})^T W_j (A_{jj}\underline{\hat{q}}_{ij} + A_{js}\underline{\hat{q}}_{is}) \\ \text{tr}(N^{-1}N_i) & \approx \bar{\gamma}_i \cdot \underline{z}_i^T G_i^T W_i (A_{ii}\underline{\hat{q}}_{ii} + A_{is}\underline{\hat{q}}_{is}). \end{array} \tag{5.21}$$

The trace $\text{tr}(N^{-1}N_i N^{-1}N_j)$ can be re-written as

$$\begin{aligned} \text{tr}(N^{-1}N_i N^{-1}N_j) &\approx \bar{\gamma}_i \cdot \left(\underline{\hat{q}}_{ij}^T N_{jj} \underline{\hat{q}}_{ij} + \underline{\hat{q}}_{ij}^T N_{js} \underline{\hat{q}}_{is} + \underline{\hat{q}}_{is}^T N_{js}^T \underline{\hat{q}}_{ij} + \underline{\hat{q}}_{is}^T A_{js}^T W_j A_{js} \underline{\hat{q}}_{is} \right) \\ &= \bar{\gamma}_i \cdot \left(-\underline{\hat{q}}_{is}^T N_{js}^T N_{jj}^{-1} N_{js} \underline{\hat{q}}_{is} + \underline{\hat{q}}_{is}^T A_{js}^T W_j A_{js} \underline{\hat{q}}_{is} \right) \\ &= \bar{\gamma}_i \cdot \underline{\hat{q}}_{is}^T \Upsilon_j \underline{\hat{q}}_{is} \\ &= \frac{\bar{\gamma}_i}{\bar{\gamma}_j} \cdot \underline{\hat{q}}_{is}^T \Upsilon_j^* \underline{\hat{q}}_{is} \end{aligned} \tag{5.22}$$

for any $i \neq j$. We use the unweighted reduced normal matrices Υ_j^* as these matrices will be stored and re-used throughout the estimations of the vector of unknowns and the VCE. The trace $\text{tr}(N^{-1}N_i N^{-1}N_i)$ can be re-written as

$$\text{tr}(N^{-1}N_i N^{-1}N_i) \approx \underline{\hat{q}}_{is}^T \Upsilon_i^* \underline{\hat{q}}_{is} + 2\underline{z}_i^T G_i^T C_i^{-1} A_{ii} \underline{\hat{q}}_{ii}. \tag{5.23}$$

The equations for MINQUE can now be expressed as

$$S\underline{\hat{\gamma}} = \underline{u} \tag{5.24}$$

with

$$\begin{array}{lll} S_{jj} & = & \frac{1}{\bar{\gamma}_j^2} \cdot \left(m_j - 2\underline{z}_j G_j^T C_j^{-1} A_{js} \underline{\hat{q}}_{js} + \underline{\hat{q}}_{js}^T \Upsilon_j^* \underline{\hat{q}}_{js} \right) \\ S_{ij} & = & \frac{1}{\bar{\gamma}_i^2} \underline{\hat{q}}_{js}^T \Upsilon_i^* \underline{\hat{q}}_{js} \qquad \text{if } i \neq j \\ u_j & = & \frac{1}{\bar{\gamma}_j^2} \underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j. \end{array}$$

Hence, we can use the scheme of figure 5.2 to estimate the vectors $\underline{\hat{q}}_i$ and then use Eq. (5.24) to compute the variance components.

5.3 Monte Carlo IREML (MC-IREML)

We will now try to implement Monte Carlo algorithms to estimate the Iterated Maximum Likelihood Estimator (IREML). The general form (Eq. (4.75)), i.e.,

$$\hat{\gamma}_i = \frac{\underline{y}^T R Q_i R \underline{y}}{\text{tr}(RQ_i)} \bar{\gamma}_i$$

will not be treated as it includes $R = Q_y^{-1} P_A^{\perp}$, which is in general not available for large-scale problems if one does not consider uncorrelated observation groups.

5.3.1 Groups with multiple variance components

Functional model of type I

The IREML for uncorrelated observation groups with multiple variance components per observation group reads (Eq. (4.102)):

$$\hat{\sigma}_{v,l}^2 = \frac{\underline{\hat{e}}_v^T W_v \bar{\sigma}_{v,l}^2 C_{vl} W_v \underline{\hat{e}}_v}{\text{tr}(W_v C_{vl}) - \text{tr}(N^{-1} A_v^T W_v C_{vl} W_v A_v)}.$$

The second trace in the denominator needs to be estimated with STE:

$$\begin{aligned} \text{tr}(N^{-1} A_v^T W_v C_{vl} W_v A_v) &= \text{tr}(N^{-1} A_v^T W_v G_{vl} G_{vl}^T W_v A_v) \\ &= \underline{z}_v^T G_{vl}^T W_v A_v N^{-1} A_v^T W_v G_{vl} \underline{z} \\ &= \underline{z}_v^T G_{vl}^T W_v A_v \underline{\hat{s}}_{vl} \end{aligned} \tag{5.25}$$

with

$$N \underline{\hat{s}}_{vl} = A_v^T W_v G_{vl} \underline{z}. \tag{5.26}$$

The observation vector $\underline{y}$ will be replaced in the black box by

$$\underline{y} = \begin{bmatrix} \underline{y}_1 \\ \vdots \\ \underline{y}_{v-1} \\ \underline{y}_v \\ \underline{y}_{v+1} \\ \vdots \\ \underline{y}_p \end{bmatrix} \rightarrow \underline{y}_{vl}^s = \begin{bmatrix} \underline{0} \\ \vdots \\ \underline{0} \\ G_{vl} \underline{z}_v \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}. \tag{5.27}$$

Note that the variance-covariance matrix of $\underline{y}_{vl}^s$ is the Q_{vl}-matrix, defined in Eq. (4.84), but again we will use the full variance-covariance matrix of the observations $\underline{y}$ as its stochastic model. The algorithm can be found in figure 5.3.

In the initial phase, we compute the factorization of C_{vl} into $G_{vl} G_{vl}^T$ and we derive the pseudo-observations $\underline{y}_{vl}^s$. These quantities have to be computed once and will not change during the iteration process.

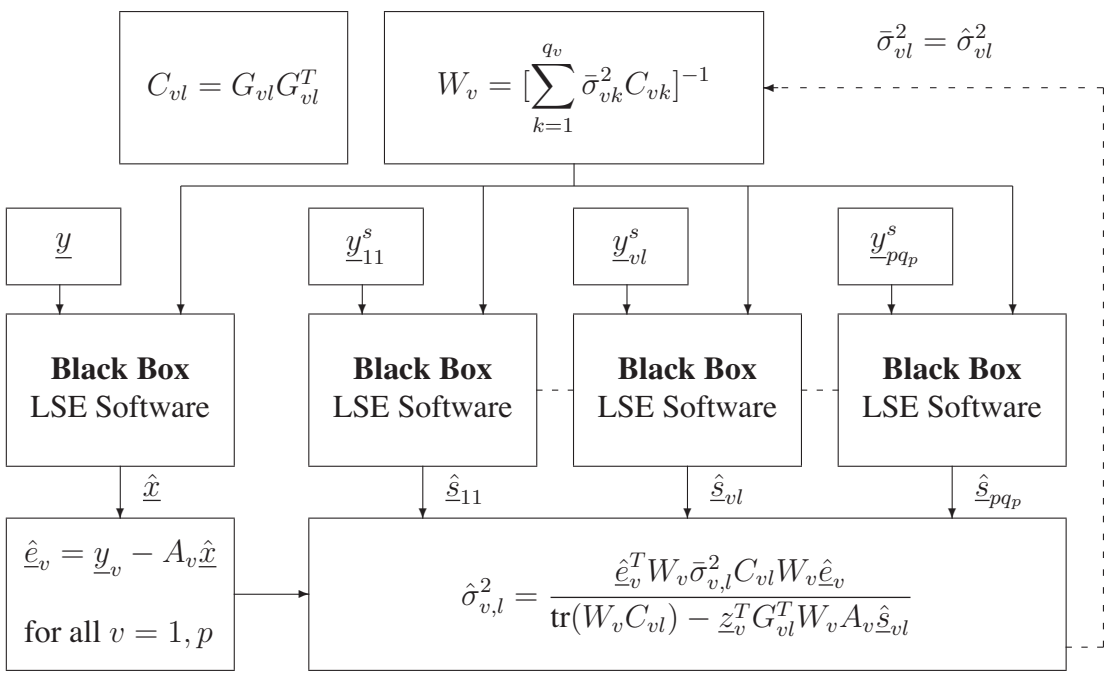


Fig. 5.3: The algorithm to estimate the vector of unknowns $\hat{\underline{x}}$ and the variance components $\hat{\underline{\gamma}}$ using IREML with uncorrelated observation groups, multiple variance components per group, and a functional model of type I.

Functional model of type II

If one considers local parameters for certain observation groups, the IREML reads (Eq. (4.105)):

$$\hat{\sigma}^2_{v,l} = \frac{\underline{\hat{e}}_v^T W_v \bar{\sigma}^2_{v,l} C_{vl} W_v \underline{\hat{e}}_v}{\text{tr}(W_v P^{\perp}_{A_{vv}} C_{vl}) - \text{tr}(\bar{N}_{ss}^{-1} A_{vs}^T W_v P^{\perp}_{A_{vv}} C_{vl} W_v A_{vs})}. \tag{5.28}$$

As in previous cases, it is easier to modify the algorithm of figure 5.3 than to estimate the traces of Eq. (4.105) by STE. Such a modification is the replacement of A_v by $[\ 0 \quad \dots \quad 0 \quad A_{vv} \quad 0 \quad \dots \quad 0 \quad A_{vs}\ ]$.

5.3.2 Groups with only one variance component

Functional model of type I

The IREML in case of one variance component per observation group reads (Eq. (4.110)):

$$\hat{\gamma}_j = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{m_j - \text{tr}(N^{-1} N_j)}.$$

The STE of this trace has been derived in Eq. (5.12) and could also be found in Kusche [2003]. The algorithm for this estimator can be found in figure 5.4.

Functional model of type II

The equations of the IREML when one considers a functional model of type II can be written as (see Eq.(4.111)):

$$\hat{\gamma}_j = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{m_i - n_i - \text{tr}(\bar{N}_{ss}^{-1} A_{js}^T W_j P^{\perp}_{A_{jj}} A_{js})} = \frac{\underline{\hat{e}}_j^T C_j^{-1} \underline{\hat{e}}_j}{m_i - n_i - \text{tr}(\bar{N}_{ss}^{-1} \Upsilon_j)}$$

The Monte Carlo version is a modification of figure 5.4. We simply have to replace the denominator of the IREML:

$$m_j - \underline{z}_j G_j^T C_j^{-1} A_j \underline{\hat{q}}_j \quad \rightarrow \quad m_j - \underline{z}_j G_j^T C_j^{-1} [A_{jj} \underline{\hat{q}}_{jj} + A_{js} \underline{\hat{q}}_{js}] \tag{5.29}$$

One could also estimate $\text{tr}(\bar{N}_{ss}^{-1} \Upsilon_j)$ with STE, but this will be of no computational benefit if one uses an existing software package as a black box.

5.4 Monte Carlo Lerch (MC-Lerch)

In Lerch's method, the variance factors are estimated by (Eq. (4.139)):

$$v_j = \frac{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})}{E\{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})\}} = \frac{(\underline{\hat{x}}_{[j]} - \underline{\hat{x}})^T (\underline{\hat{x}}_{[j]} - \underline{\hat{x}})}{\text{tr}(N_{[j]}^{-1} - N^{-1})}.$$

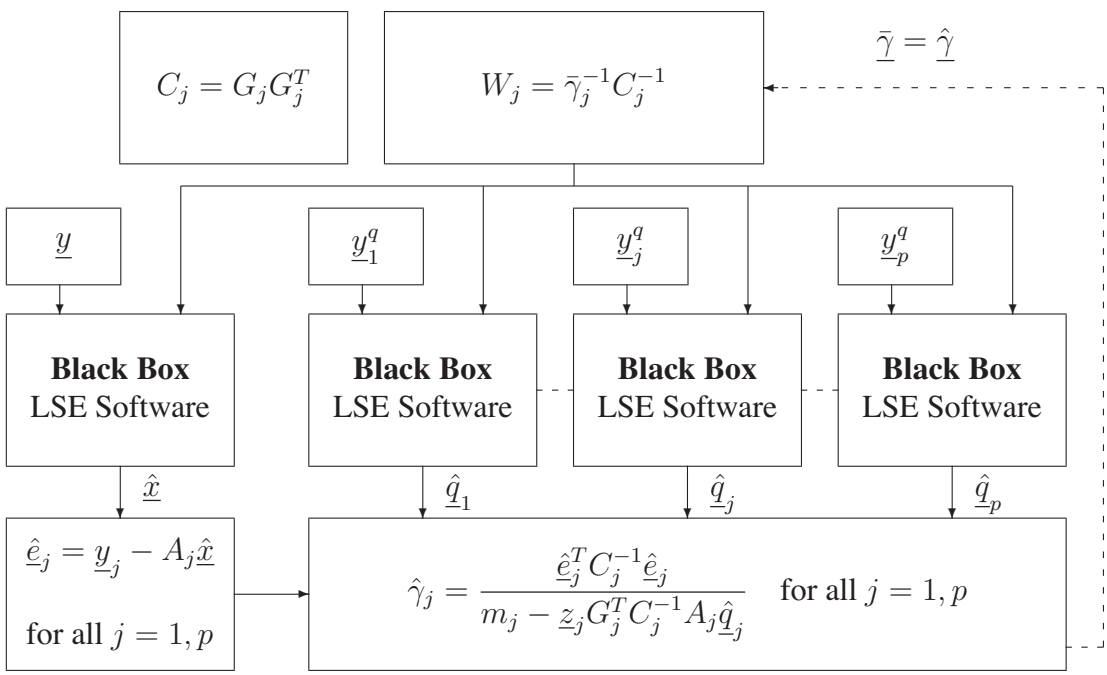


Fig. 5.4: The algorithm to estimate the vector of unknowns $\hat{\underline{x}}$ and the variance components $\hat{\underline{\gamma}}$ using IREML with uncorrelated observation groups, one variance component per group, and a functional model of type I.

The trace in the denominator can be estimated using STE techniques:

$$s_j = \text{tr}(N_{[j]}^{-1} - N^{-1}) = \underline{z}^T(N_{[j]}^{-1} - N^{-1})\underline{z}, \tag{5.30}$$

where $\underline{z}$ is a random vector with the dimension of the number of calibrated unknown parameters. The first step in the STE is to solve the systems of normal equations:

$$N_{[j]}\hat{\underline{q}}_{[j]} = \underline{z} \qquad N\hat{\underline{q}}_j = \underline{z}. \tag{5.31}$$

The solutions $\hat{\underline{q}}_{[j]}$ and $\hat{\underline{q}}_j$ are then used to estimate the denominator:

$$s_j = \underline{z}^T(\hat{\underline{q}}_{[j]} - \hat{\underline{q}}_j). \tag{5.32}$$

Note that one has to invert (or solve) a new normal matrix every time one computes a variance factor.

5.5 Summary

The computation of a high resolution, global gravity model poses a large-scale numerical problem, which usually involves millions of observations and thousands of unknown parameters. Using the equations of VCE mentioned in the previous chapters will practically not be feasible as large matrix operations must be carried out, such as the inversion of the normal matrix. This is probably the reason why variance components estimation has not widely been used in global gravity modelling.

Stochastic trace estimation (STE) is one of the solutions to this problem. The trace operator is then replaced with a pre-multiplication and post-multiplication with a random vector. As this is an approximation of the trace of the matrix, a small error is introduced. The variance of these errors attains its minimum if the entries of the random vector are samples of the binary distribution.

Both MINQUE and IREML are approximated using STE. The estimators are called Monte Carlo MINQUE (MC-MINQUE) and Monte Carlo IREML (MC-IREML), respectively. We have written the equations in such a way that if we replace the observation vector by a randomized observation vector, we can easily use any least-squares software package to obtain the estimates of the variance components. The chapter ends with a Monte Carlo variant of Lerch's subset solution method (MC-Lerch).

6

Outlier detection and robust estimation

Most observations are corrupted by small error contributions and the summation of these errors will, in general, produce normally distributed data (central limit theorem, see, e.g., Cramér [1946]). Under the assumption of normality and with *true* functional and stochastic models, the least-squares approach will produce a Maximum Likelihood estimate of the vector of unknowns $\underline{x}$, which is also a Best Linear Unbiased Estimator (BLUE), with *best* referring to minimum variance.

However, experience tells us that a set of observations is, in general, contaminated by outliers. As the least-squares method tries to minimize the square of the residuals ($\underline{e}^T Q_y^{-1} \underline{e}$), such outliers can have a significant effect on the estimated gravity field parameters; see, e.g., Götzelmann et al. [2006]. Moreover, they will affect a possible adjustment of the stochastic model using VCE.

There are several ways to detect and treat possible outliers. Detection of outliers in time series can either be done in the time domain or in the space domain. The detection in the time-domain is performed by fitting the observations to some local model, such as polynomials, splines or wavelets. A disadvantage of such a method is that it does not take into account the spatial correlation. A spatial anomaly in the field can be misinterpreted as an outlier in the time domain. Therefore, detection in the time-domain should only be performed if the residuals of the outliers are much larger than the variation in the expected observations $\underline{y}$. Outlier detection in the time domain is, however, a relative fast method and can be used to quickly detect and remove the largest outliers in a data set.

Detection in the space domain does take into account the spatial correlation of the field through the estimated vector of unknowns $\underline{\hat{x}}$, as it uses the estimated residual $\hat{e}_i$ and compares this to the estimated standard deviation of the residual $\sigma_{\hat{e}_i}$ (*Pope's test*, [Pope, 1976]), or the estimated standard deviation of the observation σ_{y_i} (μ*-test statistic*). The

least-squares estimation of the vector of unknowns should be iterated after each treatment of an outlier (*data snooping*). This is, however, not feasible in large-scale problems and alternatives will be treated in this chapter. An overview of several outlier detection techniques, which can be applied to the GOCE observables can be found in Kern et al. [2005].

Once detected, one could remove the outliers from the observation vector or reweight the corresponding observation within the least-squares approach (M-estimation). In the latter case, one assumes the observations to have a different distribution than the normal distribution.

The first part of this chapter deals with the detection of the outliers in the space domain (time domain will not be considered). The second part will focus on the reweighting of the observations using M-estimation and the consequences of such a reweighting on the VCE. Moreover, we have developed the cost function estimation (CFE) algorithm, which estimates the cost function (and consequently the probability density function) from the observations themselves.

6.1 Test statistics in outlier detection

Three different test statistics will be derived, which are commonly used in outlier detection algorithms. A discussion whether or not to include correlations when detecting outliers will be the scope of the next section. A short discussion about the treatment of the outliers is given at the end of this section.

6.1.1 The w-test statistic

We will now assume the stochastic model to be known and test for any misfit δ in a certain observation y_i. This theory is well-known and extensively described in, e.g., Teunissen [2000b]. The null hypothesis reads

$$H_0 : \underline{y} = A\underline{x} + \underline{e} \tag{6.1}$$

with

$$\underline{y} \sim N(A\underline{x}, Q_y), \tag{6.2}$$

and the alternative hypothesis can be written as (cf. the overall model test, Eq. (3.10))

$$H_A : \underline{y} = A\underline{x} + \underline{b}_i \cdot \delta + \underline{e} \tag{6.3}$$

with

$$\underline{y} \sim N(A\underline{x} + \underline{b}_i \cdot \delta, Q_y), \tag{6.4}$$

where

$$\begin{aligned} \delta &= \text{blunder in observation } i \\ \underline{b}_i &= (0 \cdots 0\, 1\, 0 \cdots 0)^T. \end{aligned}$$

Using a similar derivation as in section 3.2.2, the test statistic $\delta\hat{\Omega} = \hat{\Omega}_0 - \hat{\Omega}_A$ is distributed as

$$H_0 : \delta\hat{\Omega} \sim \chi^2(1) \quad ; H_A : \delta\hat{\Omega} \sim \chi^2(1, \lambda) \tag{6.5}$$

and can be computed by

$$\delta\hat{\Omega} = \frac{\hat{\delta}^2}{\sigma^2_{\hat{\delta}}}, \tag{6.6}$$

with

$$\hat{\delta} = \frac{\underline{b}_i^T Q_y^{-1} \underline{\hat{e}}}{\underline{b}_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} \underline{b}_i} \tag{6.7}$$

and

$$\sigma^2_{\hat{\delta}} = \frac{1}{\underline{b}_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} \underline{b}_i}. \tag{6.8}$$

The non-centrality parameter λ reads

$$\lambda = \delta^T [\underline{b}_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} \underline{b}_i] \delta. \tag{6.9}$$

The well-known w-test statistic takes the square root of the test statistic $\delta\hat{\Omega}$ [Teunissen, 2000b]:

$$\hat{w}_i = \frac{\hat{\delta}}{\sigma_{\hat{\delta}}} = \frac{\underline{b}_i^T Q_y^{-1} \underline{\hat{e}}}{\sqrt{\underline{b}_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} \underline{b}_i}}. \tag{6.10}$$

This test statistic is distributed as

$$H_0 : \hat{w}_i \sim N(0, 1) \quad ; \quad H_A : \hat{w}_i \sim N(\sqrt{\underline{b}_i^T Q_y^{-1} Q_{\hat{e}} Q_y^{-1} \underline{b}_i} \nabla, 1). \tag{6.11}$$

If we can assume uncorrelated observations, the computational effort will be greatly reduced. The test statistic $\hat{w}_i$ can then be written as

$$\hat{w}_i = \frac{\hat{e}_i}{\sigma_{\hat{e}_i}} = \frac{\hat{e}_i}{\sqrt{\sigma_i^2 - A_i N^{-1} A_i^T}}, \tag{6.12}$$

with $\sigma_{\hat{e}_i}$ the standard deviation of the residual $\hat{e}_i$, which is the square root of the ith diagonal element of $Q_{\hat{e}}$.

6.1.2 Pope's test statistic

The w-test assumes that the true stochastic model Q_y is known. If this is not true, the w-test should be replaced with the Pope's test, which assumes the stochastic model to be rescaled using VCE. The Pope's test statistic for uncorrelated observations then reads

$$\hat{\tau}_i = \frac{\hat{e}_i}{\hat{\sigma}_{\hat{e}_i}} = \frac{\hat{e}_i}{\sqrt{\hat{\sigma}_i^2 - A_i N^{-1} A_i^T}}. \tag{6.13}$$

The τ-test statistic will not have a standard normal distribution. However, with a large number of observations and a sufficiently high redundancy, this distribution will be very close to the normal distribution.

6.1.3 The μ-test statistic

The Pope's test is a good test for detecting outliers in small-scale least-squares problems. However, due to the time-consuming operations involved in this test, it is practically not feasible for large-scale problems. An alternative is to replace $\hat{\sigma}_{\hat{e}_i}$ with $\hat{\sigma}_i$, thus replacing the standard deviation of the residual by the standard deviation of the observation. In this way, we neglect $A_i N^{-1} A_i^T$. The test statistic now reads:

$$\hat{\mu}_i = \frac{\hat{e}_i}{\hat{\sigma}_i} \tag{6.14}$$

The μ-test statistic will have a slightly different distribution than the normal distribution, due to the before-mentioned approximation and the estimation of the variance components. The big advantage of the μ-test statistic is that all quantities have been computed before. Therefore the computations are extremely fast.

6.1.4 Treatment of the outliers

Under the H_0-hypothesis, the w-test statistic will have a standard normal distribution. Both Pope's test and the μ-test statistic will have slightly different distributions (under H_0), but will be close to the standard normal distribution.

A commonly used method is to remove all observations for which the chosen test statistic is above a certain threshold κ_α. With a threshold $\kappa_\alpha = 3.0$, and assuming a standard normal distribution, the one-sided level of significance will be $\alpha = 0.0013$. This threshold is widely used in a combination with the computation of the μ-test statistic. We refer to this method as the *3-sigma rule*. For other critical values, we refer to table C.1.

Due to correlations in the observations and the functional relationships, good observations could get high test statistics. It is therefore (theoretically) better to only remove the observation with the highest test statistic and then recompute $\underline{\hat{x}}$ and test again for any other outliers after each outlier removal. This is called *data snooping*; see, e.g., Baarda [1968]. However, as the method is not feasible in large-scale problems, we will remove (or down-weight) all observations of which the test statistic exceeds the threshold within a single iteration.

As the choice of the threshold κ_α is rather arbitrary and it sharply determines which observations will get full weight and which observations shall be removed, we shall discuss an alternative method in section 6.3, which will gradually give lower weights to observations with a higher test statistic.

6.2 Detection of outliers in correlated observations

There has been some discussion throughout the years, whether or not to take correlations into account when detecting outliers; see Xu [1989] and Yang et al. [2002]. The answer to this question is not as trivial as many authors suggest.

If we use the w-test statistic, we assume that one observation is an outlier (with δ the inconsistency in that observation) and all other observations are good observations, which follow the stochastic model. However, if an outlier occurs in time series, it is most likely that surrounding observations are also effected by the source of the outlier. Therefore, the residuals of these observations will not follow the assumed stochastic model. For example, if an outlier occurs in observation y_i, the observations y_{i-1} and y_{i+1} will most likely not be mutually correlated according to the stochastic model.

Furthermore, we assume that the pseudo-observation $y_i - \hat{\delta}_i$ will follow the stochastic model and also this assumption is rather premature. The source of the outlier (e.g., malfunctioning) can cause high correlations with observations *after* the blunder occurs, and negligible correlations with previous observations. An outlier is defined as an observation that does not match the assumed stochastic model, and it is questionable if the subtraction of $\hat{\delta}_i$ will change this.

A reason to use the correlations in the detection of outliers is that $\underline{b}_i^T Q_y^{-1} \underline{\hat{e}}$ (Eq. (6.10)) will de-smooth the smoothing effect of a certain outlier on correlated observations.

Therefore, in small-scale problems, we can assume correlated observations and perform data snooping. In large-size problems with large time series of observations (in which often a large number of outliers occurs), we shall neglect the correlations in the observations (only during the detection of the outliers) and remove (or down-weight) all possible outliers within a single iteration. Whether or not to treat any surrounding observations in a similar way is an interesting topic, see, e.g., Ding and Coleman [1996], who make use of the redundancy number of that observation. We choose to leave the other observations unaltered and use the VC-matrix, including the covariances, to estimate the vector of unknowns $\underline{x}$ and the variance components $\underline{\gamma}$.

Observation i can be removed by removing the observation value from the vector $\underline{y}$, deleting the ith row from the design matrix A and removing the ith row and ith column from the VC-matrix Q_y. A re-weighting can be done by only rescaling the ith row and the ith column of the matrix Q_y^{-1}. More on the subject of re-weighting can be found in the next section.

6.3 Robust estimation

A disadvantage of the outlier removal method is that the critical value κ_α decides, which observations are taken into account and which observations should be removed. There is nothing in between. An observation is either accepted with full weight or rejected and removed. Here, we consider the robust M-estimation method [Huber, 1964, 1981], in which all observations are taken into account, but can have different (smaller) weights if the residuals are large. The method is robust, as the treatment of the outliers is robust against small changes in the estimated solution vector. Moreover, one still can compute a reliable estimate of the vector of unknowns if the number of outliers is high. With such a robust method, the assumed probability of a certain large test statistic is increased, i.e.,

one assumes a different distribution than the normal distribution. Under this assumption, the estimator of the vector of unknowns does not try to fit a multivariate normal distribution to the observations, which causes high weights for the outliers. Instead, the new estimator takes into account the existence of such observations and tries to fit a better suited distribution to the observations.

We will use a simulated data set in this section, in which we have combined three datasets of varying quality into one data set. The probability density function of the standardized residuals of this data set is shown in figure 6.1. Due to the varying quality of the observations within the data set, the probability density function shows thicker tails than the standard normal distribution.

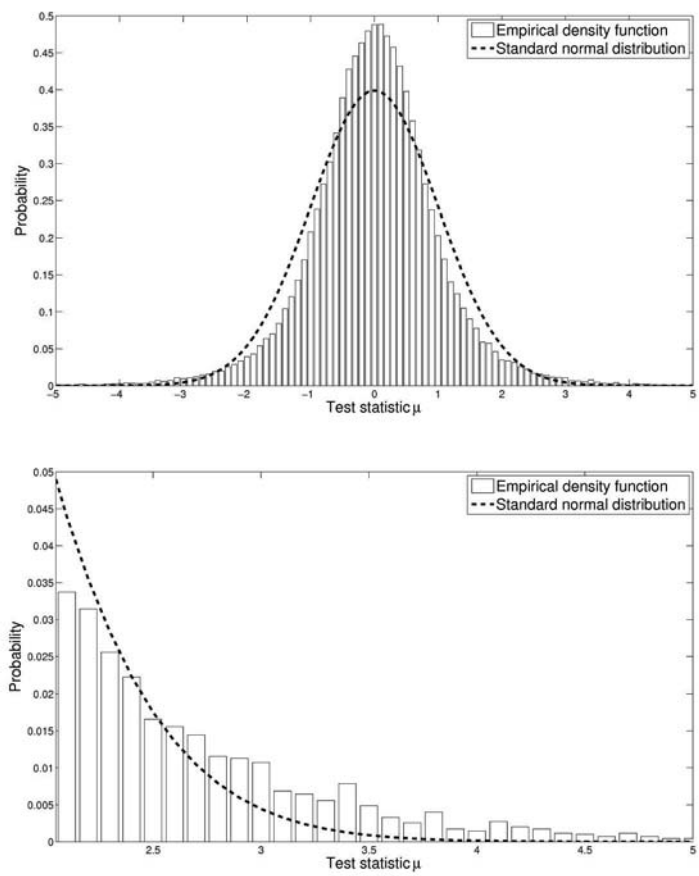


Fig. 6.1: Top: comparison between the (empirical) density function of the standardized residuals μ of a simulated data set (bars) with the density function of the standard normal distribution (dashed line). Bottom: Tail of these probability density functions.

M-estimation techniques have successfully been used with SLR data [Yang et al., 1999], magnetic Ørsted satellite data [Olsen, 2002], and in relative GPS positioning [Chang and Guo, 2005].

6.3.1 M-estimation of the vector of unknowns

The robust M-estimation assumes uncorrelated observations (in the outlier detection step) and assumes that the probability density function can be written as

$$p(\underline{y}|\underline{x}) \propto \exp[-\sum_{j=1}^{m} \rho(\mu_j)], \tag{6.15}$$

with m the number of observations and $\rho(x)$ the specific *cost function*. Maximization of the probability function $p(\underline{y}|\underline{x})$ results in the Maximum Likelihood estimate of $\underline{x}$, i.e. $\underline{\hat{x}}^R$. Such a maximization is equal to the minimization of the sum of the cost values $\rho(\mu_j)$, i.e.

$$\min_{\underline{x}} \sum_{j=1}^{m} \rho(\mu_j). \tag{6.16}$$

According to Xu [1989], the cost function $\rho(x)$ must satisfy the following:

1. Its definition domain is $\mathbb{R}$,
2. $\rho(x)$ is continuous,
3. $\rho(0) = 0$, $\rho(x) = \rho(-x)$, and if $\|a\| \leq \|b\|$ then $\rho(a) \leq \rho(b)$,
4. $\lim\limits_{x\to\infty} \dfrac{\partial^2}{\partial x^2}\rho(x) = 0.$

Minimization of the sum of cost values can also be written as

$$\sum_{j=1}^{m} \frac{\partial \rho(\mu_j)}{\partial x_i} \Big|_{\underline{x}=\underline{\hat{x}}^R} = 0 \quad \text{for all} \quad i = 1, n, \tag{6.17}$$

with n the number of unknown parameters x_i. Introducing the *influence function* $\Psi(\hat{\mu}_j)$, i.e.

$$\Psi(\hat{\mu}_j) := \frac{\partial \rho(\mu_j)}{\partial \mu_j} \Big|_{\underline{x}=\underline{\hat{x}}^R}, \tag{6.18}$$

yields

$$\sum_{j=1}^{m} \Psi(\hat{\mu}_j) \frac{\partial \mu_j}{\partial x_i} = \sum_{j=1}^{m} \Psi(\hat{\mu}_j) \frac{A_{ji}}{\hat{\sigma}_j} = 0 \quad \text{for all} \quad i = 1, n, \tag{6.19}$$

of which A_{ji} is an element of the design matrix A (Eq. (3.1)) and $\hat{\sigma}_j$ comes from Eq. (6.14). We will use the Iteratively Re-weighted Least-Squares (IRLS) method [Beaton and Tukey, 1974] to calculate the M-estimator. Defining the robust weights ν_j as

$$\nu_j^2 := \nu^2(\hat{\mu}_j) := \frac{\Psi(\hat{\mu}_j)}{\hat{\mu}_j}, \tag{6.20}$$

the n equations of the M-estimation read

$$\sum_{j=1}^{m} \frac{\nu_j^2}{\hat{\sigma}_j^2} \hat{e}_j A_{ji} = 0 \quad \text{for all} \quad i = 1, \ldots, n. \tag{6.21}$$

Writing this in matrix notation and assuming a full VC-matrix, the vector of unknowns $\underline{x}$ can be estimated by the Robust Weighted Least-Squares Estimator (RWLSE):

$$A^T D_\nu Q_y^{-1} D_\nu A \hat{\underline{x}}^R = A^T D_\nu Q_y^{-1} D_\nu \underline{y}, \tag{6.22}$$

with D_ν the diagonal matrix with the robust weights ν_i as the diagonal elements. Note that an observation is re-weighted by a factor ν_i and not by ν_i^2. One could keep the observations and design matrix fixed and change the stochastic model by re-scaling the elements of the weight matrix $W = Q_y^{-1}$:

$$W_\nu = \begin{bmatrix} \nu_1^2 w_{11} & \nu_1\nu_2 w_{12} & \cdots & \nu_1\nu_i w_{1i} & \cdots & \nu_1\nu_m w_{1m} \\ \nu_1\nu_2 w_{12} & \nu_2^2 w_{22} & \cdots & \nu_2\nu_i w_{2i} & \cdots & \nu_2\nu_m w_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \nu_1\nu_i w_{1i} & \nu_2\nu_i w_{2i} & \cdots & \nu_i^2 w_{ii} & \cdots & \nu_i\nu_m w_{im} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \nu_1\nu_m w_{1m} & \nu_2\nu_m w_{2m} & \cdots & \nu_i\nu_m w_{im} & \cdots & \nu_m^2 w_{mm} \end{bmatrix} \tag{6.23}$$

Alternatively, one keeps the weight matrix fixed and pre-multiplies the observation vector and the design matrix by the diagonal matrix D_ν. The first approach is in agreement with Yang et al. [2002]. It is an iterative method, as the weights will change with every new estimate for $\underline{x}$ and a possible new stochastic model using VCE. An alternative to the IRLS-method is Newton's method; see, e.g., Madsen and Nielsen [1990] and Chang and Guo [2005]. In general, Newton's method needs less iterations to converge than the IRLS method, as its convergence rate is quadratic, while the IRLS method has a linear rate. However, the IRLS method is easier to implement and requires less computational cost for large-size problems, in which the stochastic properties and consequently the standardized residuals change within each iteration. We therefore choose to use the IRLS method in a combination with VCE.

6.3.2 Choice of the distribution

Normal distribution

The cost function $\rho(\mu_j)$ under the assumption of a multivariate normal distribution reads

$$\rho(\mu_j) = \frac{\mu_j^2}{2}. \tag{6.24}$$

Therefore, the influence function reads

$$\Psi(\hat{\mu}_j) = \hat{\mu}_j, \tag{6.25}$$

and the 'robust' weights ν_j^2 are all equal to 1. The least-squares approach, based on this normal distribution, is, therefore, not a robust method. Outliers are equally weighted with good observations. We have to use a method, which gives less weights to observations with a higher test statistic μ_j. The most commonly used distribution within the M-estimation is Huber's distribution.

Huber's distribution

The Huber distribution [Huber, 1964, 1981] assumes the observations to be normally distributed within an interval $[-k, k]$, and a Laplace distribution outside this interval, i.e. the cost function continues as a linear function outside the interval, decreasing the influence of those observations. The distribution is defined as Eq. (6.15), where

$$\rho(\mu_j) = \begin{cases} \dfrac{\mu_j^2}{2} & \text{if} \quad |\mu_j| \leq k \\ k \cdot |\mu_j| - \dfrac{k^2}{2} & \text{if} \quad |\mu_j| \geq k. \end{cases} \tag{6.26}$$

Koch [1999] recommends that if the number of outliers is about 4% of the observations, one should use a threshold $k = 1.5$ and a threshold of $k = 2.0$ for the 1% level of outliers. Normally, the threshold is set within this interval. The influence function, which corresponds to the Huber distribution, is bounded outside the interval, i.e.,

$$\Psi(\hat{\mu}_j) = \begin{cases} \hat{\mu}_j & \text{if} \quad |\hat{\mu}_j| \leq k \\ k \cdot \hat{\mu}_j / |\hat{\mu}_j| & \text{if} \quad |\hat{\mu}_j| \geq k, \end{cases} \tag{6.27}$$

reducing the weights to

$$\nu_j^2 = \begin{cases} 1 & \text{if} \quad |\hat{\mu}_j| \leq k \\ k/|\hat{\mu}_j| & \text{if} \quad |\hat{\mu}_j| \geq k. \end{cases} \tag{6.28}$$

Tri-Weighted M-estimation

Observations with a test statistic $|\hat{\mu}_j|$ above a certain (second) threshold k_2, with k_2 ranging between 3.0 and 8.5 [Yang et al., 2005], are likely to be outliers. Huber's M-estimation down-weights these observations, instead of removing them. The Tri-Weighted M-estimation gives zero weights to observations with a test statistic outside the interval $[-k_2, k_2]$ and assumes a Huber distribution within this interval. The cost function then reads

$$\rho(\mu_j) = \begin{cases} \mu_j^2/2 & \text{if} \quad |\mu_j| \leq k_1 \\ k_1 \cdot |\mu_j| - k_1^2/2 & \text{if} \quad k_1 < |\mu_j| \leq k_2 \\ k_1 \cdot k_2 - k_1^2/2 & \text{if} \quad k_2 < |\mu_j|. \end{cases} \tag{6.29}$$

The influence function is equal to the influence function of the Huber distribution within the interval $[-k_2, k_2]$ and is zero outside this interval:

$$\Psi(\hat{\mu}_j) = \begin{cases} \hat{\mu}_j & \text{if} \quad |\hat{\mu}_j| \le k_1 \\ k_1 \cdot \dfrac{\hat{\mu}_j}{|\hat{\mu}_j|} & \text{if} \quad k_1 < |\hat{\mu}_j| \le k_2 \\ 0 & \text{if} \quad k_2 < |\hat{\mu}_j|. \end{cases} \tag{6.30}$$

The weights of the Tri-Weighted M-estimation are

$$\nu_j^2 = \begin{cases} 1 & \text{if} \quad |\hat{\mu}_j| \le k_1 \\ \dfrac{k_1}{|\hat{\mu}_j|} & \text{if} \quad k_1 < |\hat{\mu}_j| \le k_2 \\ 0 & \text{if} \quad k_2 < |\hat{\mu}_j|. \end{cases} \tag{6.31}$$

In Olsen [2002], the Tri-Weighted M-estimation was used to re-weight magnetic data from the Ørsted satellite ($k_1 = 1.5$ and $k_2 = 5.0$).

IGG-3 scheme (Yang)

A short-coming of the Tri-Weighted M-estimation is the discontinuity of the influence function at the threshold k_2. In Yang [1994] a modification of the Tri-Weighted M-estimation is proposed, in which the influence function decays towards 0, making the influence function continuous

$$\Psi(\hat{\mu}_j) = \begin{cases} \hat{\mu}_j & \text{if} \quad |\hat{\mu}_j| \le k_1 \\ k_1 \cdot \dfrac{\hat{\mu}_j}{|\hat{\mu}_j|} \cdot \left(\dfrac{k_2 - |\hat{\mu}_j|}{k_2 - k_1} \right)^2 & \text{if} \quad k_1 < |\hat{\mu}_j| \le k_2 \\ 0 & \text{if} \quad k_2 < |\hat{\mu}_j|. \end{cases} \tag{6.32}$$

with equivalent weights

$$\nu_j^2 = \begin{cases} 1 & \text{if} \quad |\hat{\mu}_j| \le k_1 \\ \dfrac{k_1}{|\hat{\mu}_j|} \cdot \left(\dfrac{k_2 - |\hat{\mu}_j|}{k_2 - k_1} \right)^2 & \text{if} \quad k_1 < |\hat{\mu}_j| \le k_2 \\ 0 & \text{if} \quad k_2 < |\hat{\mu}_j|. \end{cases} \tag{6.33}$$

Modified Yang distribution

A modification of the IGG-3 scheme is to replace the power 2 by an arbitrary power α:

$$\Psi(\hat{\mu}_j) = \begin{cases} \hat{\mu}_j & \text{if} \quad |\hat{\mu}_j| \le k_1 \\ k_1 \cdot \dfrac{\hat{\mu}_j}{|\hat{\mu}_j|} \cdot \left(\dfrac{k_2 - |\hat{\mu}_j|}{k_2 - k_1} \right)^\alpha & \text{if} \quad k_1 < |\hat{\mu}_j| \le k_2 \\ 0 & \text{if} \quad k_2 < |\hat{\mu}_j|. \end{cases} \tag{6.34}$$

With this, $\alpha = 0$ represents the Tri-Weighted M-estimation, $\alpha = 1$ corresponds to a parabolic cost function within the interval $|\mu_j| \in [k_1, k_2]$ and $\alpha = 2$ represents the IGG scheme. The parameter α can be considered as an extra free parameter, to be chosen depending on the data.

6.3.3 Cost Function Estimation (CFE)

Up till now, we have assumed that the probability density function of the observations is known. The function is either a fixed distribution, like the normal distribution or one needs to define certain shape parameters, e.g., the threshold k in case of a Huber distribution. The treatment of the outliers and consequently the estimation of the vector of unknowns and the stochastic model is, therefore, largely dependent on the rather arbitrary choice of these shape parameters.

In general, we do not have any prior information on the shape of the probability density function. However, we can use the large redundancy in the time series of satellite gravity data to estimate this probability density a-posteriori. With a sufficient parametrization of this function, this information can be used in the next iteration of the M-estimation. In this way, an estimate of the probability density function is obtained at convergence. Consequently, the parametrization of this function will produce robust weights, which better suit the error characteristics of the data than any fixed distribution.

The histogram $f'(\mu)$ of the a-posteriori standardized residuals μ_i will be a suitable estimate of the probability density function $f(\mu)$. In Constable [1988], it was suggested to fit an analytical function $\rho(\mu)$ through the empirical cost function $\rho'(\mu) = -\ln f'(\mu)$. This approach has been used in the robust weighting of geomagnetic data.

We have to find an analytical function to fit the empirical cost function. Constable [1988] suggests to use B-spline basis functions to parameterize this function. More information on the fitting of these splines can be found in Constable and Parker [1988].

Our approach is to group the bins of the empirical function $\rho'(\mu)$ and estimate a second order polynomial $\rho_i(\mu) = a_i \cdot \mu^2 + b_i \cdot \mu + c_i$ for each group of bins. If one assumes symmetry in μ_i, one could take the absolute value $|\mu_i|$ of these standardized residuals. In figure 6.2, we have made 4 groups of 10 bins to estimate the analytical cost functions $\rho_i(\mu)$ of the simulated data set of figure 6.1

With n second order polynomials ρ_i to be estimated, we should compute a total of $3 \cdot n$ coefficients. However, we have put some extra constraints on the estimation of the polynomials:

- It is not necessary for polynomials to be continuous at the interval boundaries d_i. However, the first derivatives need to be continuous at these boundaries to ensure a continuous influence function and a continuous weight function. This leads to the fixed constraint $2 \cdot a_i \cdot d_i + b_i = 2 \cdot a_{i+1} \cdot d_i + b_{i+1}$.

- The first derivative of the first polynomial must be equal to zero at $\mu = 0$. Therefore, $b_1 = 0$.

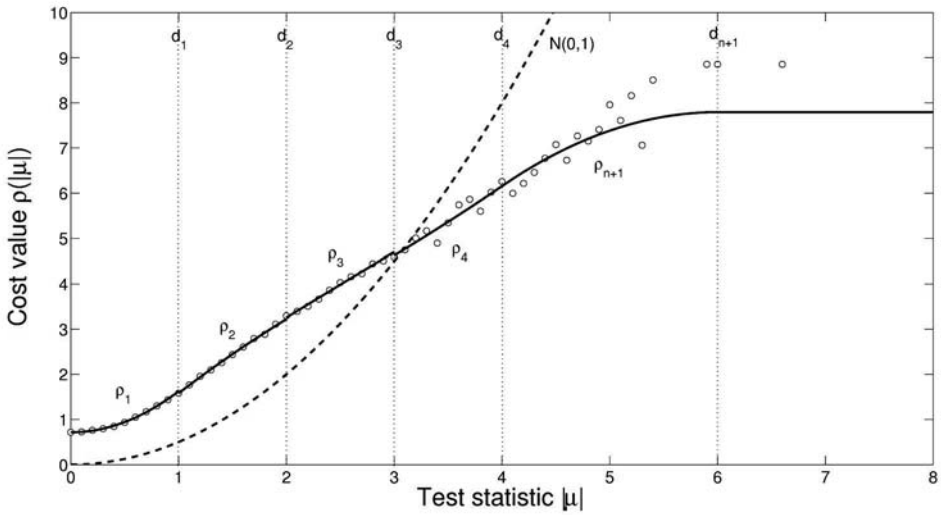


Fig. 6.2: Empirical cost function from simulated data (circles), compared to the analytical cost functions $\rho_i(\mu)$ (solid line) and the cost function of the normal distribution (dashed line).

- The last polynomial ends at d_n and is then extended to d_{n+1}. The first derivative in d_{n+1} is zero to ensure zero weight at d_{n+1}.
- The observations with a test statistic above d_{n+1} should either be removed or given zero weight.

If the fixed constraints on the parameters are written as the linear combination

$$K\underline{x} = \underline{c} \tag{6.35}$$

the constraint system of normal equations will read

$$\begin{bmatrix} N & K^T \\ K & 0 \end{bmatrix} \begin{bmatrix} \hat{\underline{x}} \\ \hat{\underline{\lambda}} \end{bmatrix} = \begin{bmatrix} \underline{h} \\ \underline{c} \end{bmatrix} \tag{6.36}$$

(see Koch [1999]), with $\underline{\lambda}$ the vector of Lagrange multipliers. Using Eq. (6.18), we can easily obtain the analytical influence function:

$$\Psi_i(\mu) = 2 \cdot a_i\mu + b_i. \tag{6.37}$$

The influence function of the simulated data set is plotted in figure 6.3.

The last part $[d_n d_{n+1}]$ of such an estimated influence function is always a straight line towards zero. The robust weights can easily be retrieved using Eq. (6.20):

$$\nu_i(\mu) = \sqrt{a_i + \frac{b_i}{\mu}}. \tag{6.38}$$

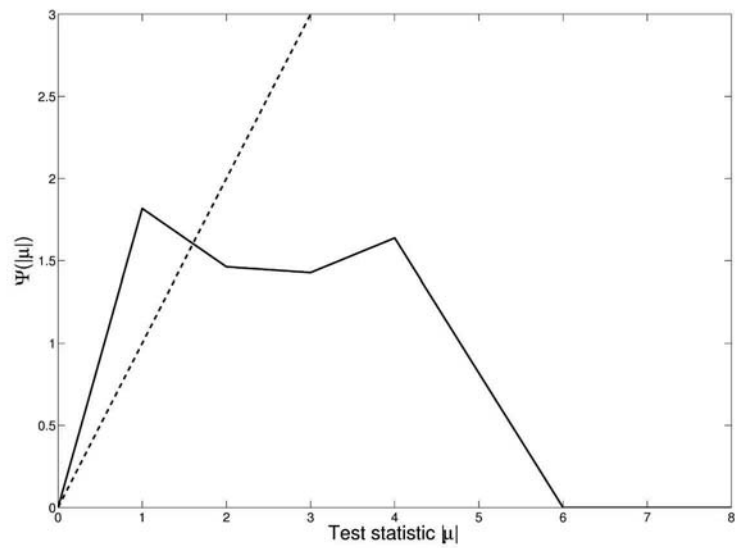


Fig. 6.3: Influence function (CFE) of the simulated data set (solid line), compared to the influence function, derived from the standard normal distribution (dashed line).

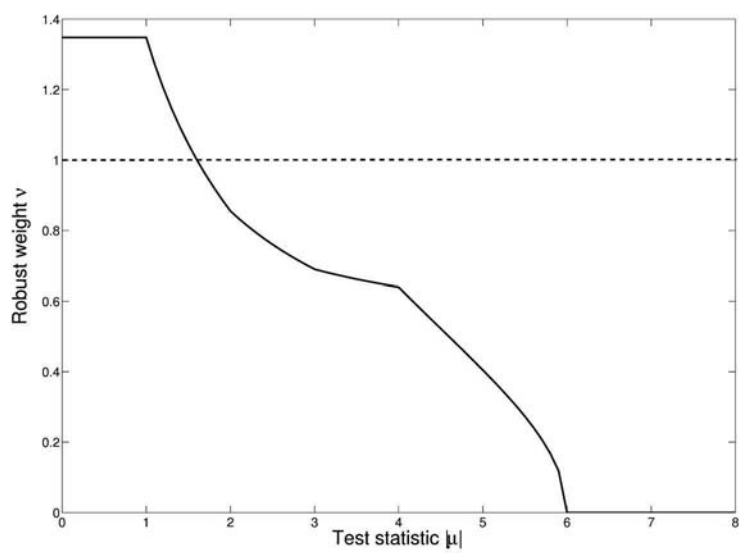


Fig. 6.4: Analytical weight function $\nu_i(\mu)$ from CFE for the simulated data set (solid) line, compared to the normal distribution (dashed line).

The robust weights of the simulated data set are shown in figure 6.4.

One could normalize the weight function by dividing it by $\sqrt{a_1}$, i.e., setting $\nu_1(\mu)$ to 1.0. However, this is not necessary, as these relative weights will be factorized later in the M-estimation of the variance components to ensure unbiasedness.

6.3.4 M-estimation of the variance components

Outliers will have a large impact on the stochastic model validation method (e.g., VCE), as the large residuals of the outliers are squared in the estimation of the variance components. Therefore, one does not only have to reduce their weight in the estimation of the unknowns, but also in the estimation of the variance components. The conventional 3-sigma rule removes all possible outliers from the data. In this way, the VCE is assumed to be free of outliers. We will focus here on the estimation of the variance components, in combination with the robust M-estimation.

Restricted Maximum Likelihood Estimation

First, we will try to derive the equations of the (Restricted) Maximum Likelihood, similar to what has been done in section (4.3.3) for normally distributed data. The REML maximizes the conditional density function

$$p(\underline{y}|\underline{\gamma}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(\underline{y}|\underline{x},\underline{\gamma}) p(\underline{x}|\underline{y}) \mathrm{d}\underline{x}. \tag{6.39}$$

The density function $p(\underline{y}|\underline{x},\underline{\gamma})$ can be written as the product of the density functions of each observation y_j, i.e.,

$$p(\underline{y}|\underline{x},\underline{\gamma}) = \prod_{j=1}^{m} p(y_j|\underline{x},\gamma_j). \tag{6.40}$$

For many M-estimators, the shape of the density function $p(y_j|\underline{x},\gamma_j)$ depends on the test statistic μ_j and consequently on y_j. This is an undesirable situation, as one needs to integrate a function over $\underline{x}$ and the shape of the function is unknown. It is therefore not possible to derive the REML of the variance components unless the shape of the density function is independent of the actual test statistic μ_j [Stahel and Welsh, 1997], like in the normal distribution or the Cauchy-Lorenz distribution (i.e. $p(x) \propto 1/(1+x^2)$).

We shall now discuss two alternative methods, which will modify the IREML estimator for normally distributed data by reducing the influence of possible outliers.

Re-weighting of the observations

The equations for the estimators of the variance components can be robustified by a re-weighting of the observations. We choose not to change the stochastic model (and weight matrix W), but to down-weight the observations by a pre-multiplication of $\underline{y}$ and

A with the diagonal matrix D_ν (see Eq. (6.22)), similar to what has been done in the robust estimation of the vector of unknowns $\underline{x}$. A scaling by a factor κ_j is necessary to make the estimator unbiased.

We shall now derive the robustified equations for several Variance Components Estimators. We start with the general case, in which we assume that all observations are correlated to each other. The equations for MINQUE (Eq. (4.28)) will change to

$$E\{\underline{u}^R\} = S^R\underline{\gamma}^R \tag{6.41}$$

with

$$\begin{aligned} S^R_{ij} &= \text{tr}(R^R Q_i R^R Q_j) \\ u^R_j &= \kappa_j \cdot (\underline{\hat{e}}^R)^T Q_y^{-1} Q_j Q_y^{-1} \underline{\hat{e}}^R, \end{aligned}$$

in which the robustified Rao's matrix R_R is defined as

$$R^R := Q_y^{-1} - Q_y^{-1} D_\nu A (A^T D_\nu Q_y^{-1} D_\nu A)^{-1} A^T D_\nu Q_y^{-1} \tag{6.42}$$

and the robustified residual vector $\underline{\hat{e}}^R$ can be computed by

$$\underline{\hat{e}}^R := D_\nu \cdot (\underline{y} - A\underline{\hat{x}}^R), \tag{6.43}$$

with $\underline{\hat{x}}^R$ the M-estimator of $\underline{x}$. If we do not take the correction factor κ_j into account, the solution will be biased, as a data set with a standard normal distribution (e.g., $\underline{z} \sim N(0,1)$) will have an estimated standard deviation of less than 1. The reason is the re-weighting of the residuals $\underline{\hat{e}}$, which will also be done for observations that follow the standard normal distribution, but have a large test statistic.

The correction factor κ_j can be computed as follows. If we write u_j as

$$\begin{aligned} u_j &= \kappa_j \cdot \underline{\hat{e}}^T D_\nu Q_y^{-1} Q_j Q_y^{-1} D_\nu \underline{\hat{e}} \\ &= \kappa_j \cdot \underline{\hat{e}}^T D_{\hat{\sigma}}^{-1} D_{\hat{\sigma}} D_\nu Q_y^{-1} Q_j Q_y^{-1} D_\nu D_{\hat{\sigma}} D_{\hat{\sigma}}^{-1} \underline{\hat{e}} \\ &= \kappa_j \cdot \underline{\hat{\mu}}^T D_{\hat{\sigma}} D_\nu Q_y^{-1} Q_j Q_y^{-1} D_\nu D_{\hat{\sigma}} \underline{\hat{\mu}}, \end{aligned} \tag{6.44}$$

with $D_{\hat{\sigma}}$ the diagonal matrix containing the diagonal elements of Q_y, the factor κ_j is derived as

$$\kappa_j = E\left\{ \frac{\underline{z}^T D_{\hat{\sigma}} Q_y^{-1} Q_j Q_y^{-1} D_{\hat{\sigma}} \underline{z}}{\underline{z}^T D_{\hat{\sigma}} D_\nu Q_y^{-1} Q_j Q_y^{-1} D_\nu D_{\hat{\sigma}} \underline{z}} \right\}, \tag{6.45}$$

and can be computed using several realizations of the $m \times 1$ vector $\underline{z}$, of which the elements have a standard normal distribution. Similar to this, we can derive the equations for the IREML in the general case, with correlated observations, i.e.

$$\hat{\gamma}_i = \frac{\kappa_i \cdot (\underline{\hat{e}}^R)^T Q_y^{-1} Q_i Q_y^{-1} \underline{\hat{e}}^R}{\text{tr}(R^R Q_i)} \bar{\gamma}_i. \tag{6.46}$$

If we assume uncorrelated observation groups with multiple variance components to be estimated per observation group, the implementation of the M-estimation algorithm to MINQUE reads (cf. Eq. (4.94))

$$E\{\underline{u}^R\} = S^R\underline{\gamma} \tag{6.47}$$

with

$$\begin{aligned} S_{ij}^R|_{u=v} &= \text{tr}(W_v C_{vk} W_v C_{vl}) - 2\text{tr}((N^R)^{-1} A_v^T D_v W_v C_{vk} W_v C_{vl} W_v D_v A_v) + \\ &\quad + \text{tr}((N^R)^{-1} A_v^T D_v W_v C_{vk} W_v D_v A_v (N^R)^{-1} A_v^T D_v W_v C_{vl} W_v D_v A_v) \\ S_{ij}^R|_{u\neq v} &= \text{tr}((N^R)^{-1} A_u^T D_u W_u C_{uk} W_u D_u A_u (N^R)^{-1} A_v^T D_v W_v C_{vl} W_v D_v A_v) \\ u_j^R &= \kappa_{vl} \cdot \underline{\hat{e}}_v^T D_v W_v C_{vl} W_v D_v \underline{\hat{e}}_v, \end{aligned}$$

where D_u is the diagonal matrix with the robust weights of data set u as diagonal elements and N^R defined as

$$N^R := A^T D_\nu Q_y^{-1} D_\nu A. \tag{6.48}$$

The corresponding robustified IREML can now be derived and reads (cf. Eq. (4.102))

$$\hat{\sigma}_{v,l}^2 = \frac{\kappa_{vl} \cdot \underline{\hat{e}}_v^T D_v W_v \bar{\sigma}_{v,l}^2 C_{vl} W_v D_v \underline{\hat{e}}_v}{\text{tr}(W_v C_{vl}) - \text{tr}((N^R)^{-1} A_v^T D_v W_v C_{vl} W_v D_v A_v)}. \tag{6.49}$$

For comparison with Fellner's approach [Fellner, 1986], we will further simplify the equations by assuming only one variance component to be estimated per observation group and the identity matrix as the cofactor matrix. The equations for the IREML simplify to

$$\begin{aligned} \hat{\gamma}_j^R &= \frac{\kappa \cdot \underline{\hat{e}}_j^T D_{\nu,j} D_{\nu,j} \underline{\hat{e}}_j}{m_j - \text{tr}((N^R)^{-1} N_j^R)} \\ &= \frac{\kappa \cdot \bar{\gamma}_j \cdot \Psi(\underline{\hat{\mu}}_j)^T \underline{\hat{\mu}}_j}{m_j - \text{tr}((N^R)^{-1} N_j^R)}, \end{aligned} \tag{6.50}$$

with $D_{\nu,i}$ the $m_j \times m_j$ diagonal matrix with the robust weights of the observations of $\underline{y}_j$, and the normal matrices defined as

$$N_j^R = A_j^T D_{\nu,j} D_{\nu,j} A_j \quad ; \quad N^R := \sum_{j=1}^{p} N_j^R. \tag{6.51}$$

The correction factor κ can now easily be derived:

$$\kappa = \frac{1}{E\left\{\Psi(z) \cdot z\right\}} \quad ; \quad z \sim N(0,1). \tag{6.52}$$

Fellner's approach

In the approach suggested by Fellner [1986] and Dueck and Lohr [2005] the residual $\hat{e}_i$ is replaced by the bounded residual

$$\hat{e}_i^B = \Psi(\hat{\mu}_i) \cdot \hat{\sigma}_i = \hat{e}_i \cdot \nu_i \cdot \nu_i \tag{6.53}$$

under the assumption that the stochastic model consists of uncorrelated observation groups with the identity matrix as cofactor matrix of group i. The normal matrices are

computed during the M-estimation of the vector of unknowns. Fellner's robust variance component can then be estimated by

$$\begin{aligned}\hat{\gamma}_i^B &= \frac{\kappa' \cdot \bar{\gamma}_i \cdot \Psi(\hat{\underline{\mu}}_i)^T \Psi(\hat{\underline{\mu}}_i)}{m_i - \mathrm{tr}((N^R)^{-1} N_i^R)} \\ &= \frac{\kappa' \cdot \hat{\underline{e}}_i^T D_{\nu_i} D_{\nu_i} D_{\nu_i} D_{\nu_i} \hat{\underline{e}}_i}{m_i - \mathrm{tr}((N^R)^{-1} N_i^R)},\end{aligned} \tag{6.54}$$

with

$$\kappa' = \frac{1}{D\{\Psi(z)\}} \quad ; \quad z \sim N(0,1). \tag{6.55}$$

Hence, the re-weighting of the observations is not consistent in Fellner's approach. The observations, and consequently the elements of the normal matrices (which enter into the denominator of Eq. (6.54)), are re-weighted by a factor ν_i, contrary to the residuals in the nominator of Eq. (6.54), which are re-weighted by a factor ν_i^2.

Decreasing influence function

One should be cautious to use Fellner's approach or the re-weighting of the observations if the influence function decreases for high test statistics μ_i. This can happen, e.g., with a cost function estimation or the tri-weighted M-estimation. The residual square sum $\Psi(\hat{\underline{\mu}}_j)^T \Psi(\hat{\underline{\mu}}_j)$, or to a lesser extend $\Psi(\hat{\underline{\mu}}_j)^T \hat{\underline{\mu}}_j$, will decrease towards zero at convergence for a data set with several outliers. However, these outliers will be counted in the number of observations (and consequently in the redundancy number), resulting in too optimistic values for the variance components. In the next iteration, the observations will be further down-weighted, which decreases the residual square sum even further. The variance component will converge to a too optimistic value, which can even be zero. A solution to this is to use a bounded influence function (e.g., Huber's influence function) in the estimation of the variance components.

6.4 Summary

In the least-squares approach, one tries to minimize the residual square sum of the observations. However, a set of observations is in general contaminated by outliers. These outliers can have large effects on the estimation of the vector of unknowns and the variance components.

The outliers should therefore be detected and effectively be down-weighted or removed. As outliers are observations, which do not follow the assumed stochastic model, one should be careful in assuming correlations between the observations in the detection step of the outliers. We suggest to consider correlations in the least-squares estimation of the vector of unknowns and in the VCE, but to neglect correlations in the detection of outliers, when dealing with large time series of satellite gravity data.

Hypothesis testing to detect an outlier in the set of observations leads to the classical w-test. An approximation is to use the standardized residual, i.e., the residual (after least-squares adjustment) divided by the a-posteriori standard deviation of the observation, which can be estimated by VCE.

Conventional outlier removal techniques remove those outliers that have a standardized residual (test statistic) above a certain threshold. An observation is good or bad and there is nothing in between. Robust M-estimation looks at the outlier treatment problem from a different point of view. It assumes a different distribution than the normal distribution and therefore assumes a higher probability for the outliers. In this way, possible outliers are down-weighted in the least-squares estimation.

We suggest to go one step further and estimate the probability density function using the adjusted residuals, i.e. we perform a *Cost Function Estimation (CFE)* on the standardized residuals. This can only be done if a sufficiently large number of residuals is available. The estimation of the cost function depends on the estimated stochastic model and vice versa. Therefore the variance component estimation, the estimation of the parameters of the probability density function and the outlier detection (and down-weighting) should be done in an iterative process.

A robustification of the VCE was suggested by Fellner [1986]. The residuals are bounded to have less influence on the estimation of the variance components. We found that this method down-weights the outliers twice, compared to the down-weighting of the observations in the estimation of the vector of unknowns. We therefore suggest to modify Fellner's approach and down-weight the observations to an equal amount for both VCE and the estimation of the vector of unknowns.

7

Application 1: CHAMP satellite gravity data

In this chapter, we will derive pseudo-observations from the primary data collected on board the CHAMP satellite using the energy balance approach (EBA). The different arcs of the CHAMP data are re-weighted using Monte Carlo variants of MINQUE, IREML and Lerch's subset solution method. Different outlier treatment algorithms are compared, such as the conventional 3-sigma rule, M-estimation and the Cost Function Estimation. A quality description of the derived satellite-only global gravity models is obtained by taking the difference with the EIGEN-GL04C model [Förste et al., 2008], which is much more accurate in all bandwidths than any CHAMP-only model. The best satellite-only solution is then combined with the EGM96 model. Inconsistencies between the two models are tested and, where necessary, estimated and a re-estimation of the stochastic properties of the EGM96-model is performed. Part of the results have been published in Van Loon [2007], and early results on the combination with prior information are published in Van Loon and Kusche [2007].

We have used CHAMP data to test the different algorithms and to be able to compare the methods to better, independent, global data from the GRACE/LAGEOS combined model EIGEN-GL04C. However, these algorithms can also be applied to any other time series, e.g., to data obtained from the GRACE or GOCE satellite missions.

7.1 The energy balance approach for gravity recovery

The energy balance approach is based on the fact that the motion of a satellite orbiting the Earth is highly correlated with the gravity field of the Earth. Other forces, such as tidal forces or non-conservative forces, will either be measured or modelled and their contribution will be removed from the data. The total energy of gravity (plus modelled forces) and kinematics is conserved. The basic idea of the energy balance approach

goes back to the early days of satellite geodesy [O'Keefe, 1957]. In Jekeli [2001], it was shown that, when neglecting some smaller effects (e.g., electromagnetic interaction, polar motion, precession, nutation), the gravity field of the Earth can be modelled with the use of the satellite's orbit vector, the knowledge of the non-conservative forces, the Earth's rotation and models concerning the tides and loading effects. For any position ($\underline{r}$) of the satellite, with t the point in time, the gravitational potential at that point can be estimated by Jekeli [1999]:

$$V(\underline{r}) \approx \frac{1}{2}\|\underline{\dot{r}}\|^2 - \int_{t_0}^{t} \underline{f}(\tau)\cdot\underline{\dot{r}}(\tau)\mathrm{d}\tau - \omega_e(r_1\dot{r}_2 - r_2\dot{r}_1) - \int_{t_0}^{t} \nabla V_{tides}(\tau)\cdot\underline{\dot{r}}(\tau)\mathrm{d}\tau - C, \quad (7.1)$$

with

$$\begin{array}{ll} V(\underline{r}) & = \text{gravitational potential at position } \underline{r}\ [\mathrm{m^2 s^{-2}}], \\ \underline{\dot{r}} & = \text{time-derivative of the position vector } \underline{r}\ [\mathrm{ms^{-1}}], \\ \underline{f} & = \text{vector of non-conservative forces } [\mathrm{ms^{-2}}], \\ \omega_e & = \text{rotation of the Earth } [\mathrm{rad/s}], \\ V_{tides} & = \text{model of the tidal potential } [\mathrm{m^2 s^{-2}}], \\ C & = \text{energy constant } [\mathrm{m^2 s^{-2}}]. \end{array}$$

This form of the EBA is often referred as the Jacobi integral. Alternative energy balance relations can be found in Löcher and Ilk [2007]. The EBA has been successfully used in global gravity field modelling using CHAMP orbits (e.g., Han et al. [2002]), and GRACE orbits (e.g., Han et al. [2005b]).

The velocity of the satellite is measured by hl-SST GPS measurements. For each arc, i.e., piece of orbit with no data gaps in the GPS measurements, the pseudo-observations $\underline{\nu}_i$ of Eq. (7.1) can be computed by:

$$\nu_k = \frac{1}{2}\|\underline{\dot{r}}_k\|^2 - \Delta t \sum_{j=0}^{k} \underline{f}_j \cdot \underline{\dot{r}}_j - \omega_e(r_{k,1}\dot{r}_{k,2} - r_{k,2}\dot{r}_{k,1}) - \Delta t \sum_{j=0}^{k} \nabla V_{tides,j} \cdot \underline{\dot{r}}_j - C_i, \quad (7.2)$$

with Δt the time interval between two consecutive GPS measurements and C_i the energy constant for arc i. The velocity vector $\underline{\dot{r}}_k$ can be derived from the orbit positions $\underline{r}_k$ using numerical differentiation. The vector of non-conservative forces $\underline{f}_j$ is measured by the on board accelerometer. The rotation of the Earth ω_e is assumed to be known and is mainly derived from VLBI measurements. Tidal models are used to estimate V_{tides}.

Many research groups (Han et al. [2002], Howe and Tscherning [2003], Kang et al. [2003] and Perosanz et al. [2003]) have found a bias, a drift and a scaling factor in the measurements of the accelerometer. Together with the energy constant, these low-frequency errors in the pseudo-observations need to be subtracted from the observations by, e.g., using a low-degree polynomial. As the coefficients of this polynomial are different for each arc, these coefficients will be treated as local parameters in the adjustment procedure. To estimate these parameters, one should first subtract an a-priori global gravity field $\underline{x}_R$ from the observations:

$$\underline{y}_i = \underline{\nu}_i - A_{is}\underline{x}_R \quad (7.3)$$

Next, a low-degree polynomial will be fitted through these (reduced) pseudo-observations to account for the low-frequency noise. The functional model then reads (cf. Eq. (3.19)):

$$\begin{aligned} E\{\underline{y}_i\} &= E\{\underline{\nu}_i\} - A_{is}\underline{x}_R \\ &= A_{is}\underline{x}_G + A_{ii}\underline{x}_{ii} - A_{is}\underline{x}_R \\ &= A_{is}\underline{x}_s + A_{ii}\underline{x}_{ii} \end{aligned} \tag{7.4}$$

with $\underline{x}_G$ the unknown spherical harmonic coefficients of the Earth's gravity field.

Other research teams remove the local parameters (drift, polynomial, bias) in a pre-processing step, based on ν_k time series and / or cross-over adjustments. Yet we consider them as local parameters in a joint estimation with the global spherical harmonic coefficients. One of the advantages of this approach is that we can study the correlations of the local parameters with the spherical harmonic coefficients.

7.2 Test setup 1: One data set (232 days)

Available data

Two years of CHAMP kinematic orbit data have been kindly provided by D. Švehla, IAPG, TU Munich. These orbits were processed following the zero–differencing strategy [Švehla and Rothacher, 2003], and were provided with full three-dimensional variance-covariance information per data point. Whenever the standard deviation exceeds the threshold of $10\ \mathrm{cm}$, the data point is removed and the data batch is cut into two parts. In this way, the two year data set is split into 9476 batches.

We start with the 1250 longest data batches, ranging between 2.7 hours and 20.4 hours. The total number of selected observations is 668465, equivalent to 232 days.

To derive kinematic velocities from the kinematic orbit we use the 7-point Lagrange interpolator:

$$\dot{x}_i \approx \frac{a_3 \cdot x_{i+3} + a_2 \cdot x_{i+2} + a_1 \cdot x_{i+1} - a_1 \cdot x_{i-1} - a_2 \cdot x_{i-2} - a_3 x_{i-3}}{c \cdot \Delta t}, \tag{7.5}$$

in which $a_1 = 45$, $a_2 = -9$, $a_3 = 1$ and $c = 60$. The accelerometer and quaternion data have been provided by GFZ, Potsdam. The direct tidal acceleration is modelled by JPL DE ephemeris, the solid earth tides follow IERS conventions and we use GOT 99.2 to account for the ocean tides [Ray, 1999]. The derived pseudo-observations are assumed to be uncorrelated.

Estimated global parameters

The gravitational field of the Earth, V, is a harmonic function outside the sources and can, therefore, be written as

$$V(\vartheta, \lambda, r) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} [\bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda] \bar{P}_{lm}(\cos\vartheta) \tag{7.6}$$

As it is not possible to estimate an infinite number of coefficients, a global model of the gravitational field is usually a truncated version of this harmonic expansion:

$$V_G(\vartheta,\lambda,r) = \frac{GM}{R}\sum_{l=0}^{L}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}[\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda]\bar{P}_{lm}(\cos\vartheta). \quad (7.7)$$

As it was stated before, to remove the low-degree polynomial, one needs to subtract a reference model from the observations. We have subtracted pseudo-observations induced by the EGM96 model [Lemoine et al., 1998]. The scalar $A_{is}\underline{x}_s$, as a part of the functional model can, therefore, be written as

$$\begin{aligned}\delta V(\vartheta_i,\lambda_i,r_i) &= A_{is}\underline{x}_s \\ &= \frac{GM}{R}\sum_{l=0}^{L}\left(\frac{R}{r_i}\right)^{l+1}\sum_{m=0}^{l}[\Delta\bar{C}_{lm}\cos m\lambda_i + \Delta\bar{S}_{lm}\sin m\lambda_i]\bar{P}_{lm}(\cos\vartheta_i),\end{aligned} \quad (7.8)$$

with $\Delta\bar{C}_{lm}$ and $\Delta\bar{S}_{lm}$ the reduced spherical harmonic coefficients. In this test setup, we truncate the global model to degree and order $L = 75$, with fixed coefficients for degrees $l = 0$ and $l = 1$, resulting in a total number of 5772 coefficients. The truncation degree $L = 75$ is chosen to minimize the sum of commission errors (noise + aliasing) and omission errors. However, as the high degrees ($50 \leq l \leq 75$) are poorly resolved by CHAMP and they suffer from spatial aliasing of the higher degrees ($l > 75$), we will only compare the solution up to degree $l = 50$.

The EIGEN-GL04C model

The EIGEN-GL04C global gravity model (figure 1.1) is a combined model, complete to degree 360, derived from GRACE data, LAGEOS data, and surface gravity data [Förste et al., 2008]. The LAGEOS data mainly improve the low-degree spherical harmonics (up to degree 30), the GRACE data are used up to degree 150, and the surface gravity data (gravimetric and altimetric) strengthen the entire bandwidth, up to degree and order 360. As the model does not include CHAMP satellite data, the model can be used to evaluate the CHAMP satellite-only solutions. The combination models computed in this chapter (a combination with the EGM96 model) can not be considered independent of the EIGEN-GL04C model, as some surface gravity data is used in both models. The mean accuracy (commission error) up to degree 360 is estimated to be 15 cm. The North-South pattern seen in many GRACE-only models is largely reduced. We will use this combination model to evaluate our satellite-only and combination models.

Estimated local parameters

We have estimated (and tested) 6 local parameters per data set, i.e., a bias and a drift in the observations, and parameters to account for two periodic effects (1 cycle per revolution (cpr) and 2 cpr). The local parameters were tested for significance (see section

(3.2.2)), and the test results can be found in table 7.1. As such tests can only be performed with a proper stochastic model, we have computed the test statistics after the estimation of the variance components with MINQUE.

Tab. 7.1: Test statistics Ω for the significance of the local parameters ($\kappa_\alpha = 3.842,\ \alpha = 0.05$), with t the time of the observation and T the period of revolution.

Parameter	% of data sets ($\Omega < \kappa_\alpha$)	Statistics of Ω		
		mean	min.	max.
Bias	$0.00\,\%$	$2.67 \cdot 10^{12}$	$8.28 \cdot 10^{9}$	$1.46 \cdot 10^{13}$
Drift	$7.44\,\%$	$1.79 \cdot 10^{3}$	$9.91 \cdot 10^{-5}$	$1.38 \cdot 10^{5}$
$\cos(2\pi \cdot t/T)$	$54.00\,\%$	10.79	$4.77 \cdot 10^{-5}$	$1.28 \cdot 10^{3}$
$\sin(2\pi \cdot t/T)$	$50.32\,\%$	11.02	$3.37 \cdot 10^{-6}$	$6.30 \cdot 10^{2}$
$\cos(4\pi \cdot t/T)$	$37.04\,\%$	20.20	$1.58 \cdot 10^{-4}$	$3.50 \cdot 10^{2}$
$\sin(4\pi \cdot t/T)$	$37.60\,\%$	23.31	$8.80 \cdot 10^{-6}$	$3.75 \cdot 10^{2}$

The test statistics for the bias and drift in the observations clearly show that these two parameters are significant, as the majority of the data sets have a test statistic Ω above the critical value κ_α. The bias absorbs the energy constant of the EBA and the drift absorbs the bias in the accelerometer, the latter under the assumption that the along-track satellite velocity is nearly constant. The four parameters of the periodic effects can also be considered significant, although in approximately half of the data sets they were found to be insignificant. As the mean and maximum values of these test statistics are well above the critical value κ_α, we can conclude that also these 4 local parameters are significant and should be included in the least-squares estimation procedure.

As we solve for spherical harmonic coefficients and local parameters in a joint estimation process, we can compute correlations between these parameters. In Kusche and Van Loon [2005], we have shown that the correlations between the local parameters and the spherical harmonic coefficients are negligible. In this study, we have computed the mean correlations among the local parameters. Apart from strong correlations between the bias and the drift, we can see that the mean correlations all lie within the interval $[-0.20, 0.20]$. The correlation between the bias and drift is not only increasing for smaller data sets [Kusche and Van Loon, 2005], but it also shows a periodic function at the orbit frequency and twice the orbit frequency; see figure 7.1. One should therefore avoid over-parametrization in the case of very small data sets, as the local parameters will be strongly correlated.

7.3 Re-weighting of the data sets

Equal weighting

We start with the most simple experiment, in which we weight all 1250 data sets equally and we do not try to find any outliers in the data. This is the standard approach in

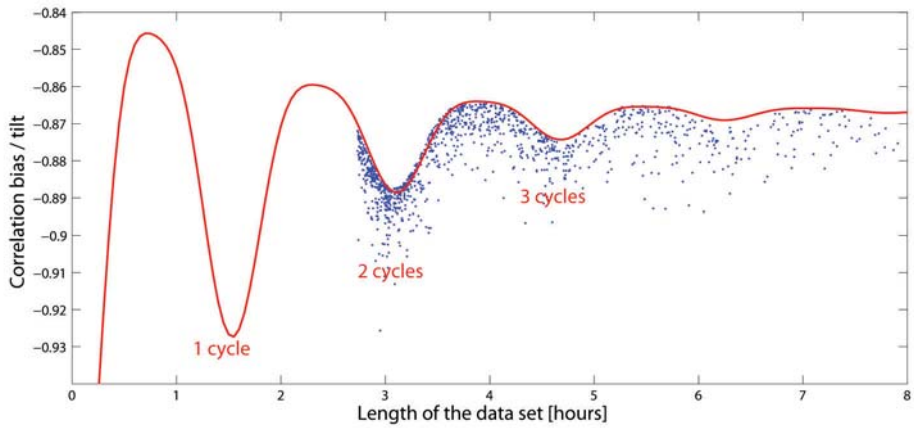
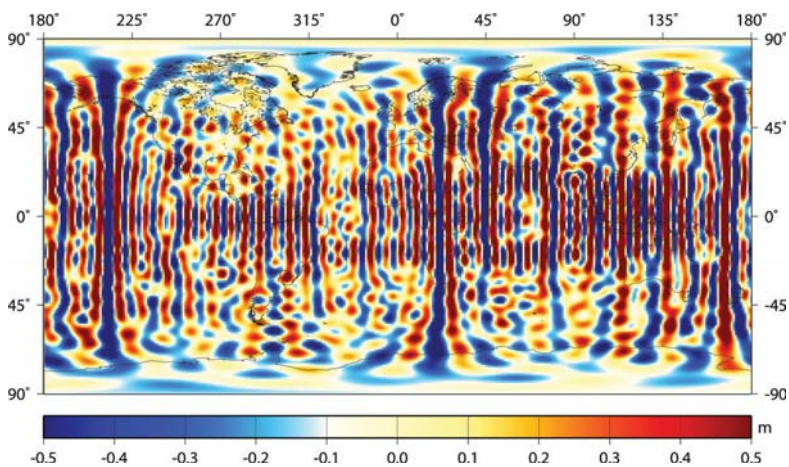


Fig. 7.1: Correlations between the estimated bias and drift for 1250 data sets (blue), with in red an approximative function.

satellite gravity field modelling using the newest CHAMP or GRACE data. We will show that the solution can be improved significantly using VCE techniques and a proper treatment of the outliers. A plot of the geoid height differences with EIGEN-GL04C (up to degree 50) can be found in figure 7.2. The solution contains large errors and a clear North-South pattern is visible in the geoid height differences with EIGEN-GL04C. The erroneous data should be down-weighted by VCE.



Error statistics [m]

min.	-2.071
max.	2.104
mean	0.000
RMS	0.420

Fig. 7.2: Geoid differences [m] between the satellite-only solution and EIGEN-GL04C ($l \leq 50$) before re-weighting of the data sets and without a treatment of the outliers.

MINQUE

As we assume normally distributed observations, the MINQUE estimator will have minimum variance (BIQUE) and converges to maximum likelihood (REML-fs). Moreover, as the data sets are assumed to be uncorrelated, the Helmert VCE will produce the same results. We have estimated the variance components using a Monte Carlo variant of MINQUE. Only three iterations were necessary to obtain convergence, as can be seen in figure 7.3.

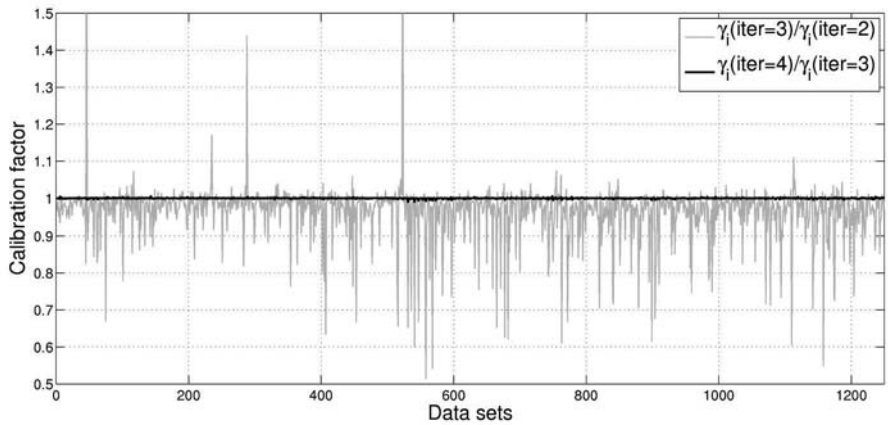


Fig. 7.3: Calibration factor between iteration 2 and 3 and between iteration 3 and 4

This figure shows that iteration 4 would give almost the same results as iteration 3. Hence, only 3 iterations were necessary to obtain convergence (of which the first iteration is the 'equal weight solution'). The estimated standard deviations (iteration 3) for all 1250 data sets are displayed in figure 7.4.

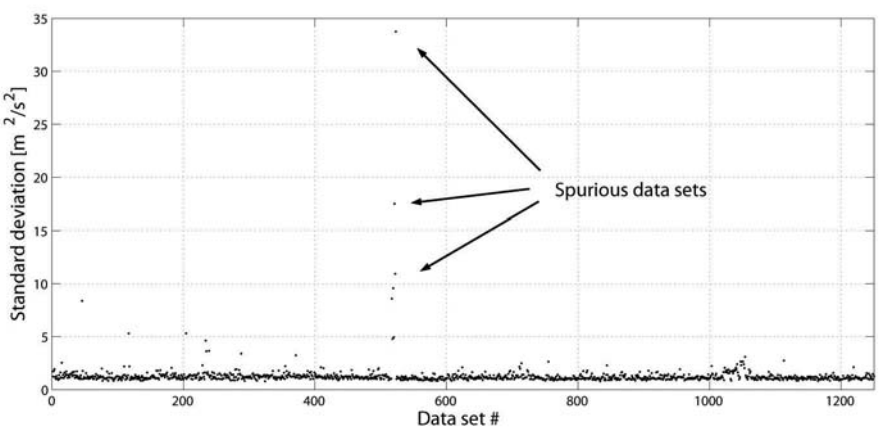


Fig. 7.4: Estimated standard deviations

It becomes clear that some spurious data sets are present in the CHAMP gravity data. These data could have caused the stripe pattern in the earlier solution using equal weights. Due to the near-polar orbit of the CHAMP satellite, some bad orbits will deteriorate the solution with some erroneous north-south pattern. The spurious data sets are however detected by MINQUE and consequently down-weighted. If one assumes the data to be normally distributed, one could estimate the uncertainty in the estimates of the variance components; see Eq. (4.33). The standard deviations (uncertainties) of the estimated variance components (BIQUE or REML-fs) are plotted as error bars in figure 7.5. The estimation of the variance components of the spurious data sets is less accurate (up to $\sigma = 18.3\,\mathrm{m}^4/\mathrm{s}^4$), compared to the other data sets, which had an accuracy of $\sigma \approx 0.06 - 0.10\,\mathrm{m}^4/\mathrm{s}^4$. The VCE does not only down-weight the spurious data sets,

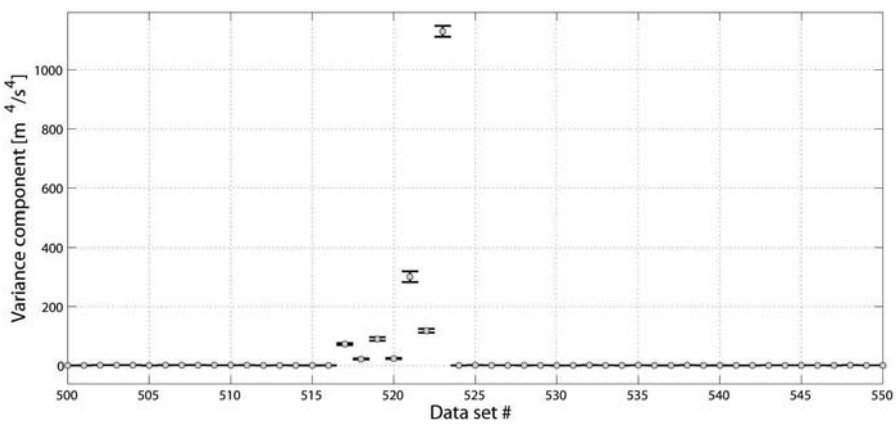


Fig. 7.5: Uncertainty in the BIQUE and REML-fs estimates for data sets 500-550 (including the spurious data sets).

but also re-weights the other data sets, of which the quality can vary considerably. The estimated standard deviations and corresponding weights can be found in figure 7.6.

The maximum likelihood of the estimated standard deviations lies close to $1.0\,\mathrm{m}^2/\mathrm{s}^2$. The figure of the weight estimates (figure 7.6, right) shows the variations of the weight estimates for each data set. If we use these weights in the Weighted Least-Squares Estimation, we can improve the solution considerably. The statistics of the geoid height differences with EIGEN-GL04C are shown in table 7.2.

The table shows the large improvements by VCE in terms of geoid height differences with EIGEN-GL04C, for different bandwidths, i.e., $l \leq 10$ and $l \leq 50$. Even if we remove the sequence of the 7 bad data sets, the VCE decreased the geoid height differences from $\sigma = 19.1\,\mathrm{cm}$ to $\sigma = 12.7\,\mathrm{cm}$ for $l \leq 50$. A spatial plot of these geoid height differences is shown in figure 7.7. The erroneous stripe pattern, present in the equal weighting solution, has vanished once a proper stochastic model has been implemented using VCE. Hence, VCE has successfully down-weighted these spurious data sets.

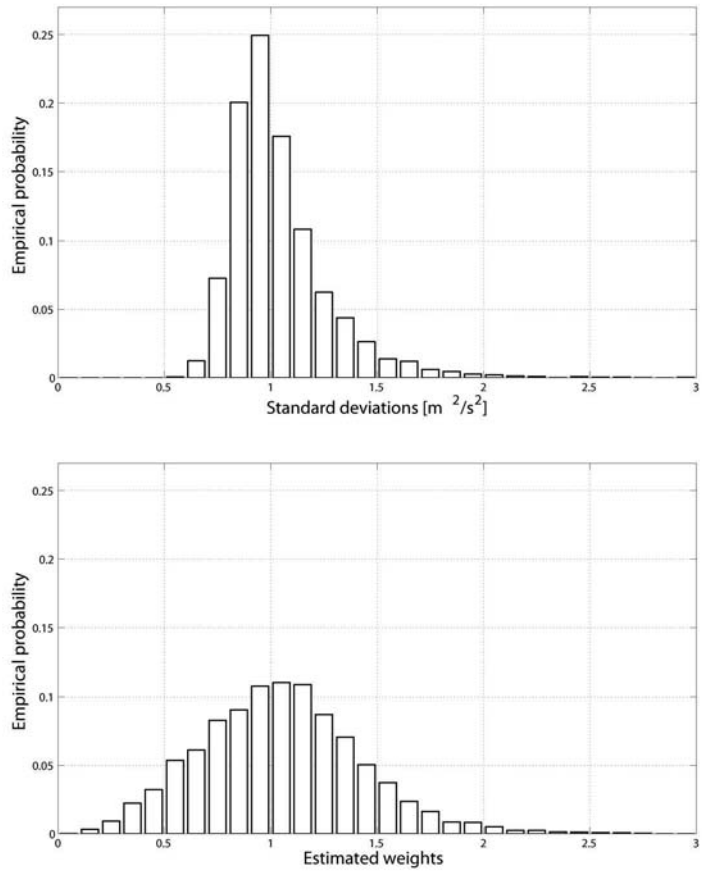
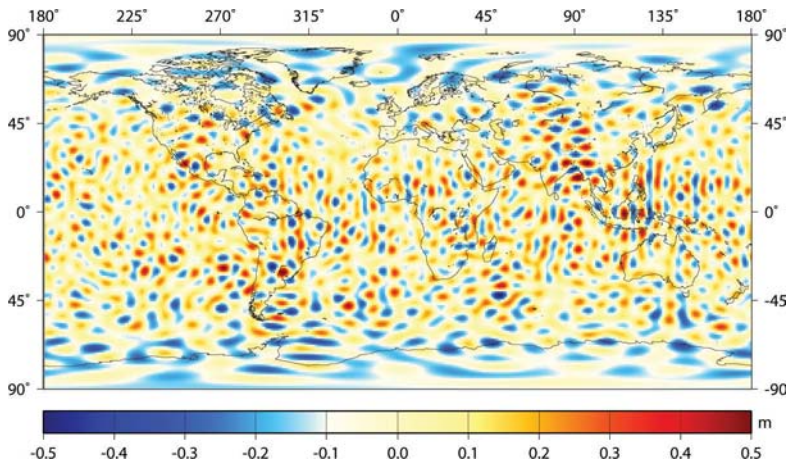


Fig. 7.6: Top: estimated standard deviations of the CHAMP data sets. Bottom: corresponding weights of these data sets.

Tab. 7.2: Statistics of the geoid height differences [m] between various satellite-only solutions and EIGEN-GL04C. The solutions differ in the way the weights of the data sets are determined.

VCE method	Data sets	$l \leq 10$ Max.	$l \leq 10$ RMS	$l \leq 50$ Max.	$l \leq 50$ RMS
$\gamma_i = 1.0$	1250	0.137	0.045	2.104	0.420
$\gamma_i = 1.0$	1243	0.101	0.036	0.920	0.191
MC-MINQUE	1250	0.104	0.036	0.616	0.127
MC-IREML	1250	0.104	0.036	0.616	0.127
MC-Lerch	1250	0.102	0.035	0.563	0.126
IMLE	1250	0.104	0.036	0.616	0.127



Error statistics [m]

min.	-0.547
max.	0.616
mean	0.000
RMS	0.127

Fig. 7.7: Geoid differences [m] between the satellite-only solution and EIGEN-GL04C ($l \leq 50$) after a re-weighting of the data sets using MINQUE (3 iterations). No treatment of the outliers.

Iterative Restricted Maximum Likelihood Estimator (IREML)

The Iterative Restricted Maximum Likelihood Estimator (IREML) converges to the same estimates as the MINQUE. The results are therefore equal to the MINQUE results; see table 7.2. However, the Monte Carlo IREML is much faster than the Monte Carlo variant of MINQUE. In this test setup, the computations of MC-MINQUE took 100 times more CPU time than the computation of the MC-IREML estimates.

The computational costs of MC-MINQUE increase with the number of independent elements of S, i.e. $p{\cdot}(p{+}1)/2$, whereas the computational costs of MC-IREML increase with p, where p is the number of variance components to be estimated. The choice whether to use MC-IREML or MC-MINQUE is therefore dependent on the size of the problem, if one requires a quality description of the estimated variance components and if negative variance components are allowed.

Lerch's subset solution method

We have also tested the Monte Carlo variant of Lerch's subset solution method. The estimated standard deviations are compared with the estimated MC-MINQUE standard deviations in figure 7.8. Although Lerch's subset solution method is derived from a different point of view than the MINQUE method, the results are very similar. The results in terms of geoid difference with EIGEN-GL04C are therefore almost identical to the MC-MINQUE results. The estimated standard deviations for the spurious data sets are however too high in Lerch's method, compared to the unbiased MINQUE estimates.

The main disadvantage of Lerch's subset solution are the computational costs, as one has to solve p systems of normal equations. In this test setup, the CPU time of the subset solution method was by a factor 155 larger than the computation of the IREML estimates.

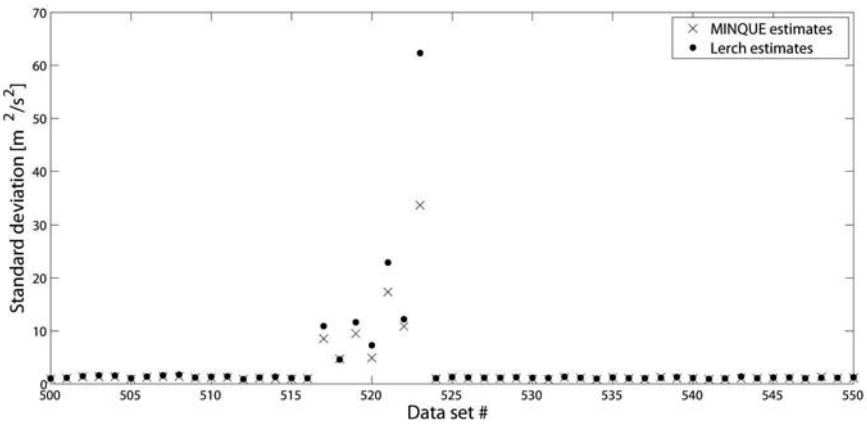


Fig. 7.8: Comparison between the estimates of Lerch's subset solution method and MINQUE estimates for data sets 500-550.

Iterative Maximum Likelihood Estimator (IMLE)

The results of the IMLE are, in this test setup, almost the same as the MINQUE estimates; see figure 7.9. The difference between IMLE and IREML is that we neglect the trace of the group influence matrices when computing the IMLE. As all groups only have a small influence on the total solution, compared to the number of observations, we could approximate the redundancy number by using the number of observations.

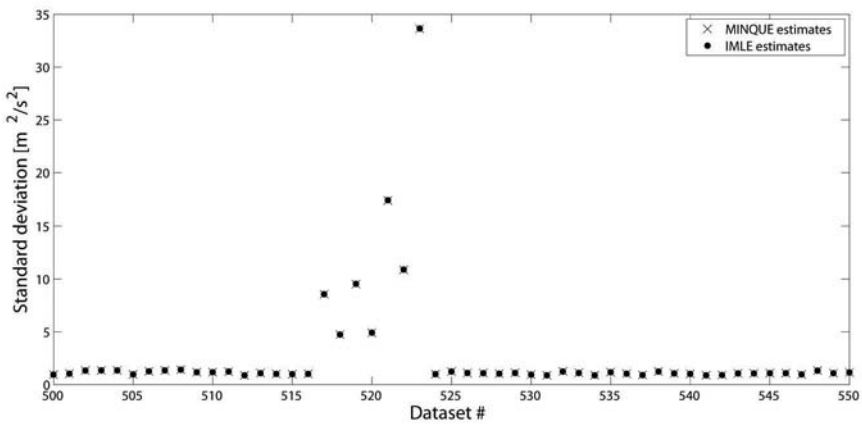


Fig. 7.9: Comparison between the IMLE and MINQUE estimates for data sets 500-550.

In this test setup, the quantities to compute the IMLE were easy to compute: $\underline{\hat{e}}_i^T \underline{\hat{e}}_i / m_i$ and the computational costs are therefore negligible compared to the other VCE algorithms. As can be concluded from table 7.2, all four proposed methods give comparable results in this CHAMP experiment. Both MC-MINQUE and MC-Lerch methods are

however quite time-consuming. As the MC-IREML method and the IMLE method are both fast and give comparable results, we choose to use the MC-IREML method in further computations, as this estimator is almost unbiased, compared to the biased IMLE method.

7.4 Treatment of outliers

The stochastic model (and consequently the weights of the data sets) will be determined by using the IREML (with Monte Carlo techniques to speed up the computations). However, to achieve reasonable variance estimates of this stochastic model, one should first remove the effect of possible outliers. These outliers will deteriorate both the solution of the vector of unknowns as the estimation of the stochastic properties of the data.

We will compare three methods to each other: the conventional 3-sigma rule, in which one removes any possible outliers, the robust M-estimation and finally the Cost Function Estimation (CFE) technique, in which one down-weights possible outliers. We start with the conventional 3-sigma rule.

Conventional 3-sigma rule

We have combined the μ-test to detect and remove outliers ($\kappa_\alpha = 3.0$) with the MC-IREML method to estimate the stochastic model. As outliers have an influence on the estimation of the variance components and can only be detected with a proper stochastic model, the two algorithms are connected, and we need several iterations for the MC-IREML and the computation of the μ-test statistic.

We have used three iterations, in which all observations with a test statistic μ above the threshold $k = 3.0$ are removed within each iteration. In this way, $2.1\ \%$ of the CHAMP pseudo-observations has been removed. The results in terms of geoid height differences with EIGEN-GL04C are shown in table 7.3. We can see that the removal of the outliers has almost no effect on the estimation of the very low degrees ($l \leq 10$), no matter which outlier treatment technique is used. However, the solution up to degree 50 improved from $12.7\ \mathrm{cm}$ to $11.6\ \mathrm{cm}$ in terms of RMS of the geoid height differences with EIGEN-GL04C.

M-estimation

The drawback of the conventional 3-sigma rule is the strict separation between good and bad observations. An observation with a test statistic of $\mu = 2.99$ is considered to be a good observation and is given full weight, whereas an observation with a test statistic of $\mu = 3.00$ is considered to be an outlier and is removed. In the robust M-estimation method, we increase the probability of outliers compared to the normal distribution. In this way, possible outlying observations are given less weight. There is no clear cut-off.

Tab. 7.3: Statistics of the geoid height differences [m] between the satellite-only solutions and EIGEN-GL04C.

VCE method	Outlier treatment	$l \leq 10$		$l \leq 50$	
		Max.	RMS	Max.	RMS
$\gamma_i = 1.0$	-	0.137	0.045	2.104	0.420
MC-IREML	-	0.104	0.036	0.616	0.127
MC-IREML	3-sigma	0.100	0.036	0.507	0.116
Fellner	Huber	0.102	0.036	0.482	0.113
Fellner	tri-weighted	0.101	0.036	0.454	0.112
Fellner	CFE	0.099	0.037	0.470	0.113
Fellner (Huber)	CFE	0.100	0.036	0.459	0.111

We have tested two well-known M-estimation weight functions: the Huber weight function [Huber, 1964] and the tri-weighted weight function [Yang et al., 2005]. Both weight functions start down-weighting the observations with a test statistic above $k_1 = 1.7$. However, outliers with a very high test statistic, e.g., above 10, still receive a substantial weight. These observations are given zero weight, when using the tri-weighted M-estimation. We have chosen to put this zero-weight threshold at $k_2 = 4.5$. Fellner's approach has been used to remove the effect of outliers in the estimation of the variance components.

The geoid height differences with respect to the EIGEN-GL04C model are given in table 7.3. The tri-weighted method gives slightly better results than the Huber distribution, as the largest outliers are given zero weight.

Cost Function Estimation (CFE)

When considering the Huber distribution or the tri-weighted weight function, one assumes that the observations have a certain fixed distribution. Based on this assumption, one assigns pre-defined (robust) weights to the observations. In the CFE approach, we use the time series of the test statistic μ_i to estimate the distribution from the observations. We consider the class of probability functions, which has the form of Eq. (6.15) and can therefore uniquely determine the probability density function from the cost function $\rho(\mu)$. The empirical cost function can now be retrieved by taking $\rho(\mu) = -\ln f(\mu)$, of which $f(\mu)$ is approximated by the histogram of the test statistics μ_i. The empirical cost function can be found in the top panel of figure 7.10, together with the estimated cost function, which is a linear combination of second order polynomials. The figure shows the cost function estimates after the three VCE iterations (iter=1) and after the second series of three VCE iterations (iter=2). As both cost function estimates (and corresponding weighting schemes) are almost identical, we can stop after two iterations of CFE.

The analytical function $\rho(\mu)$, expressed in sets of polynomials of the second order, allows us to estimate the probability density function $f(\mu)$, the influence function $\Psi(\mu)$,

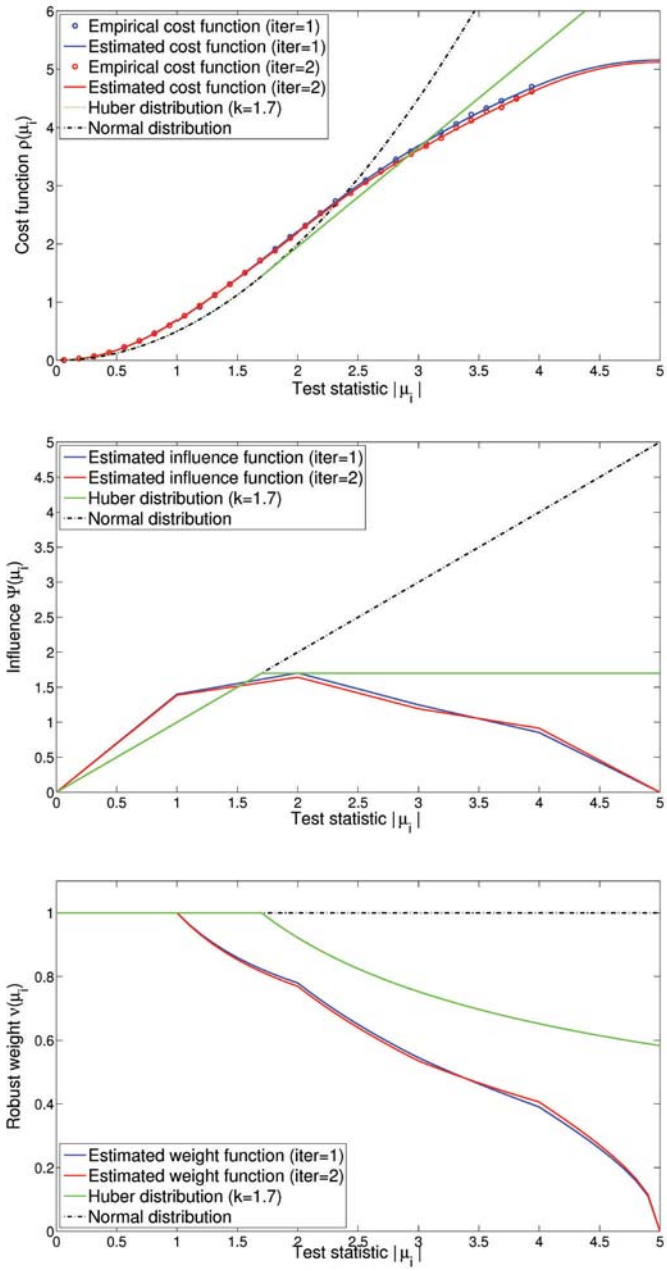


Fig. 7.10: Top: empirical cost functions from 232 days of CHAMP pseudo-observations, compared with the estimated analytical cost functions, for different iterations of CFE. The boundary parameters are $d_n = 4.0$ and $d_{n+1} = 5.0$. The interval $[0, d_n]$ is divided into $n = 4$ groups and each polynomial is estimated out of 8 bins. Middle: influence function $\Psi(\mu)$. Bottom: (normalized) weight function $\nu(\mu)$.

and the weighting function $\nu(\mu)$. The influence functions and the weight functions for iteration 1 and iteration 2 are shown in the middle and bottom panel of figure 7.10. The weight function is scaled to ensure $\nu(0) = 1.0$ for comparisons with the normal distribution and the Huber distribution. This has no effect on the estimation of the vector of unknowns, as it only changes the absolute weights and not the relative weights. Moreover, the scaling factor κ in Fellner's approach or RVCE will ensure unbiasedness in the variance components.

As the influence function $\Psi(\mu)$ decreases towards zero, we have used the Huber influence function in the computation of the variance components. This gave the best results in terms of geoid height differences with EIGEN-GL04C, compared to other M-estimation methods and the conventional 3-sigma rule method; see table 7.3. We will now perform similar tests for a larger data set, i.e., the full two-year data set.

7.5 Test setup 2: Two year data set

We will now use the tested algorithms on the full data set of two years. The data set consisted of 9476 different data batches (see section 7.2). As the correlations between the local parameters increase drastically for smaller data sets (section 7.2) and the efficiency (group influence / computational costs) decrease for smaller data sets, we excluded the smallest data batches from this test setup. The smallest data set used had a time span of 27 minutes. After the removal of the data batches lacking accelerometer data, we could group the remaining batches according to time into 7 subsets of 1228 or 1229 data batches. This was equivalent to 594 days. The number of observations for each subset of the two-year data set is shown in figure 7.4.

Tab. 7.4: Number of observations for each subset of the two-year data set

Data set	Observations	Equivalent days
1	174917	60.7
2	237364	82.4
3	174443	60.6
4	243897	84.7
5	283519	98.4
6	274759	95.4
7	320537	111.3
Total	1709436	593.6

We have used the data of data set 1 to compare the CFE to the tri-weighted weight function and to compare Fellner's approach to the re-weighting of the observations (RVCE), see section 6.3.4. The results in terms of geoid height differences with EIGEN-GL04C can be found in table 7.5.

We can conclude from table 7.5 that again the CFE performs better compared to the conventional 3-sigma rule or the M-estimation. Both Fellner's approach and RVCE

Tab. 7.5: Geoid height differences [m] between the satellite-only solutions of data set 1 and EIGEN-GL04C.

VCE method	Outlier treatment	$l \leq 10$		$l \leq 50$	
		Max.	RMS	Max.	RMS
$\gamma_i = 1.0$	-	0.218	0.072	3.032	0.472
IREML	-	0.191	0.063	2.383	0.443
IREML	3-sigma	0.164	0.048	1.963	0.404
Fellner	tri-weighted	0.174	0.050	1.922	0.405
Fellner	CFE	0.168	0.048	1.817	0.386
Fellner (Huber)	CFE	0.166	0.049	1.810	0.387
RVCE	tri-weighted	0.172	0.049	1.892	0.393
RVCE	CFE	0.169	0.049	1.813	0.387
RVCE (Huber)	CFE	0.166	0.049	1.804	0.388

show similar results. However, we prefer to use RVCE, as this method is consistent with the M-estimation of the vector of unknowns. As data set 1 does not contain arcs with many outliers, the Huberization of the VCE does not have much effect. However, we will use this Huberization for the other data sets in case we encounter a data batch with many outliers.

From each data set, a satellite-only gravity model is computed using the CFE, with an Huberized RVCE. The estimated cost functions for the 7 subsets are shown in figure 7.11.

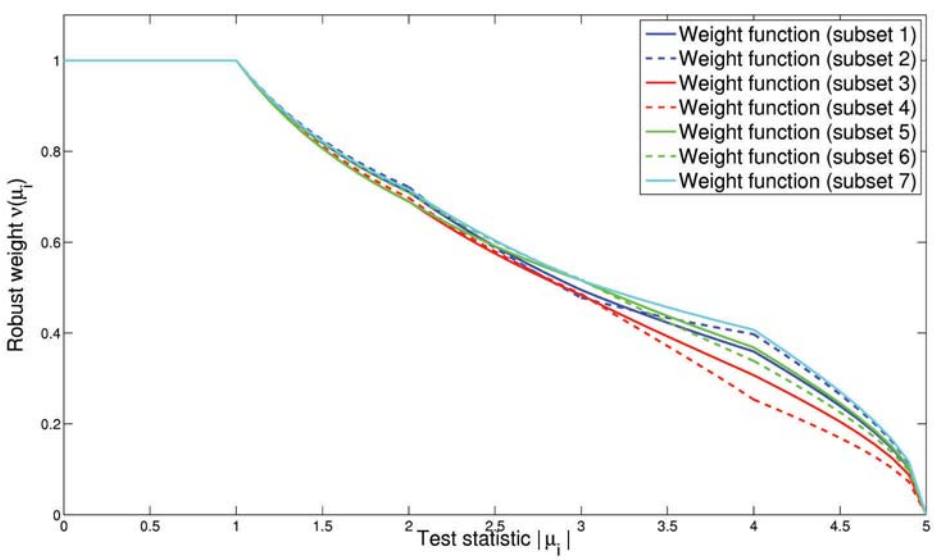


Fig. 7.11: The CFE weight functions for the 7 independent subsets.

We can conclude from this figure that the distributions of the independent subsets

of observations are consistent with each other. The cost functions only deviate from each other at the higher test statistics, as the accuracy of the polynomials for these test statistics decrease through decreasing number of observations within a bin. The statistics of the geoid height differences with EIGEN-GL04C are shown in table 7.6.

Tab. 7.6: Left: geoid height differences [m] between the 7 subset solutions and EIGEN-GL04C. Right: estimated variance components in a joint estimation (Huber, RVCE)

subset	$l \leq 10$		$l \leq 50$		IREML	RVCE
	Max.	RMS	Max.	RMS	$\hat{\gamma}_i$	$\hat{\gamma}_i$
1	0.166	0.049	1.804	0.388	3.74	3.30
2	0.101	0.035	0.892	0.190	2.74	2.01
3	0.101	0.031	1.804	0.368	3.07	2.76
4	0.080	0.027	0.955	0.215	2.70	2.43
5	0.145	0.054	1.137	0.252	3.79	3.27
6	0.154	0.060	0.758	0.169	2.53	1.99
7	0.081	0.029	0.767	0.160	2.07	1.83

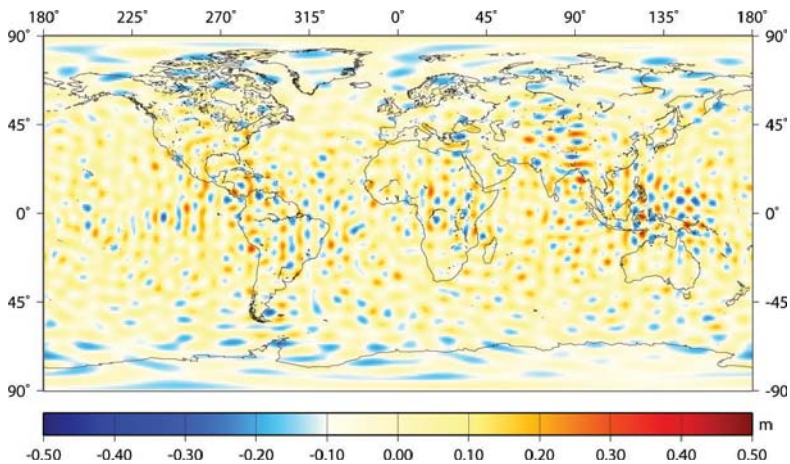
At this point, we can look at the 7 satellite-only solutions as independent observation groups with the spherical harmonic coefficients as input observations. We have combined the 7 data sets using REML to re-weight the observation groups and the Huber distribution to re-weight the spherical harmonic observations (and down-weight possible outliers). The estimated variance components for the non-robust and robust case are shown in table 7.6. It can be seen that a bad data set in terms of geoid height differences with EIGEN-GL04C (external comparison), also obtains (using VCE) a high scaling of the stochastic model (internal calibration). Moreover, one can clearly see in table 7.6 that the estimated variance components of the robustified IREML are much lower than the variance components of the non-robust IREML, due to a down-weighting of outlying spherical harmonic coefficients. The statistics of the geoid height differences of the joint solution with EIGEN-GL04C are shown in table 7.7.

Tab. 7.7: Statistics of the geoid height differences [m] between the different satellite-only solutions and EIGEN-GL04C.

VCE	Outlier	$l \leq 10$		$l \leq 50$	
method	treatment	Max.	RMS	Max.	RMS
$\gamma_i = 1.0$	-	0.096	0.036	0.387	0.088
IREML	-	0.094	0.034	0.387	0.086
RVCE	Huber	0.078	0.027	0.383	0.084

The RVCE satellite-only solution is clearly better in the very low degree ($l \leq 10$) than the solutions without a proper outlier treatment. The lowest degrees ($l \leq 4$) are

poorly resolved by satellite-only solutions and the CHAMP solution show inconsistencies for these degrees with independent models, which we will discuss in the next section. A plot of the geoid height differences with EIGEN-GL04C is shown in figure 7.12.



Error statistics [m]

min.	-0.314
max.	0.383
mean	0.000
RMS	0.084

Fig. 7.12: Geoid differences [m] between the CHAMP satellite-only solution and EIGEN-GL04C ($l \leq 50$) after a re-weighting of the data sets by RVCE.

7.6 Combination with a prior gravity model

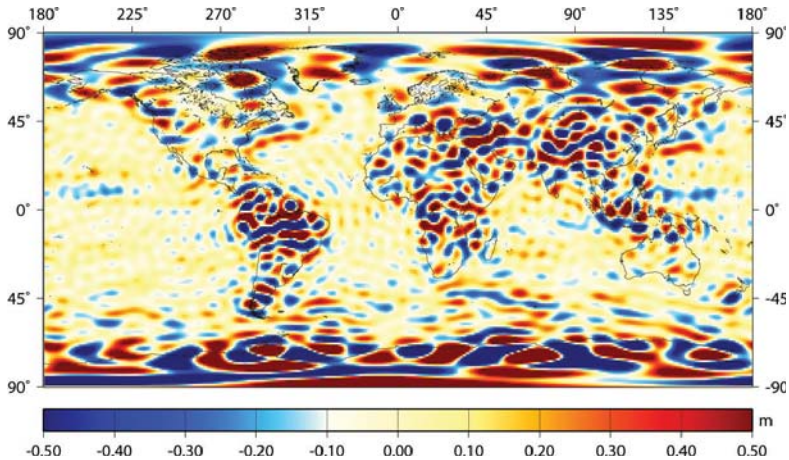
7.6.1 The EGM96 global gravity model

The EGM96 global gravity model [Lemoine et al, 1998], which is complete up to degree 360, had been the most widely used global gravity model before the launch of the dedicated satellite missions CHAMP and GRACE. The model combines SLR data with altimeter, terrestrial, marine and airborne gravimetry data. The re-weighting of the data sets was partly done using Lerch's subset solution method and partly by trial and error comparisons with independent data.

The model scores very good in the oceanic regions, due to the availability of radar altimeter data. A spatial plot of the geoid height differences with EIGEN-GL04C is shown in figure 7.13. As some surface gravity data have been used in both EIGEN-GL04C and EGM96 models, these models can not be considered independent from each other.

7.6.2 Stochastic model validation (first iteration)

If we want to combine the EGM96 model with the CHAMP satellite-only model, the two data sets should be consistent with each other, in terms of the solution vector and in the stochastic model. Any inconsistency between the two data sets, present in the



Error statistics [m]

min.	-2.230
max.	3.350
mean	0.000
RMS	0.259

Fig. 7.13: Geoid height differences [m] between EGM96 and EIGEN-GL04C ($l \leq 50$).

solution vector, should be tested and consequently removed from one of the two data sets. Such a testing can only be done properly if the stochastic models are accurate enough, both in relative terms and in absolute terms. However, the estimation of the stochastic properties of a data set, e.g., by VCE, assumes the functional model to be valid. Any systematic effects or inconsistencies should have been augmented in this functional model.

As we expect that a proper stochastic model has more effect on the test of significance of the inconsistencies than the augmentation of the functional model on the stochastic model validation, we have chosen to start with such a stochastic model validation. After this, inconsistencies are tested and if necessary dealt with by an augmentation of the functional models. Finally, the stochastic properties are re-estimated using the augmented functional models.

(Robust) re-weighting of the data sets

We have compared the robust combination of EGM96 and CHAMP with the IREML combination and the equal weight combination, of which the latter is the most common procedure in combining a satellite-only solution with prior information in global gravity field modelling. The advantage of the robust combination (Huber, $k = 1.7$) is that badly determined coefficients of either the CHAMP solution or the EGM96 solution are down-weighted and will not disturb the determination of the weights by VCE. The results are shown in table 7.8.

We can see that large improvements are made using the robust combination, compared to the other two combination methods, especially in the lowest degrees. The variance component of CHAMP was estimated at 5.79, the EGM96 variance component at 4.82. Figure 7.14 shows the geoid height differences by degree. It becomes clear that some low-degree spherical harmonic coefficients in CHAMP are badly determined

Tab. 7.8: Statistics of the geoid height differences [m] between different combination solutions and EIGEN-GL04C.

Solution		$l \leq 10$		$l \leq 50$	
		Max.	RMS	Max.	RMS
EGM96		0.062	0.020	3.350	0.259
CHAMP		0.078	0.027	0.383	0.084
EGM96 + CHAMP	$\gamma_i = 1.0$	0.031	0.010	0.393	0.072
	IREML	0.029	0.010	0.480	0.075
	RVCE	0.019	0.006	0.417	0.069

or even biased with respect to EIGEN-GL04C and EGM96. The RVCE combination down-weights such spurious spherical harmonic coefficients. In this way these coefficients are better determined in the combination solution. Moreover, it improves the estimation of the other coefficients, due to the mutual correlations, and it improves the estimates of the variance components.

In the bottom panel of figure 7.14, the focus is on the very low degree spherical harmonics. A comparison is made between the non-robust IREML solution and the robust RVCE solution. The IREML solution is close to the CHAMP-only solution due to the much better stochastic model of the CHAMP spherical harmonic coefficients. The difference between the CHAMP-only solution and the combined IREML solution, which is the residual vector of CHAMP in the first iteration of the RVCE solution, is however still too large for the low-degree coefficients, if we compare it to the re-scaled stochastic model of CHAMP. Therefore, the standardized residuals $\hat{\mu}_i$ are very high for the low-degree spherical harmonic coefficients of CHAMP. This leads to lower robust weights in the next iteration. At convergence, the robustified stochastic models fit better the actual residuals of EGM96 and CHAMP compared to the RVCE solution.

The robust RVCE combination enables us to treat the vector of unknowns (spherical harmonic coefficients) better by down-weighting outlying coefficients. In this way, inconsistent low-degree CHAMP coefficients could be down-weighted, which improved the combination solution considerably.

Eigenvalue decomposition

The publicly available VC-matrix of EGM96 is dense up to degree $l = 70$ and has a diagonal structure for higher degrees [Lemoine et al., 1998]. The VC-matrix is the outcome of a specific weighting scheme (Lerch subset solution and trial and error) of several different data sets. It would therefore be better if we were able to re-weight the different data sets ourselves with the use of VCE algorithms. As these data are not publicly available, we try to get as close as possible by using an eigenvalue decomposition (EVD) of the EGM96 VC-matrix.

By grouping the eigenvalues, we can split the VC-matrix into several cofactor matrices. These cofactor matrices are then re-scaled using a VCE. In this way, the new

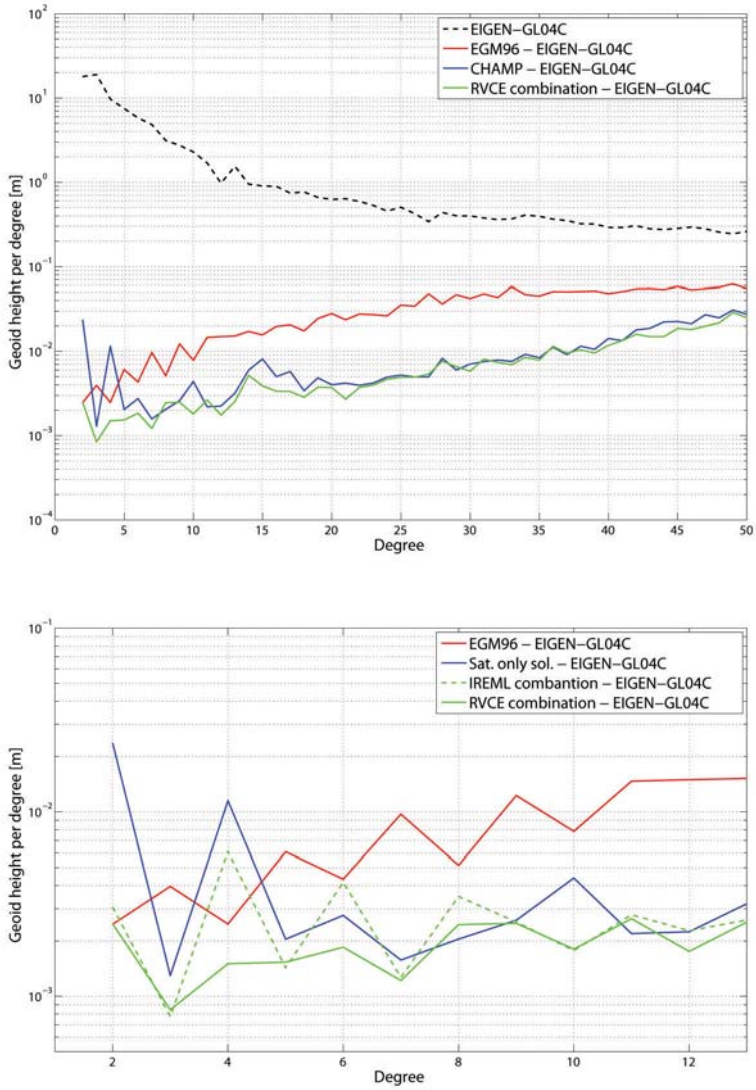


Fig. 7.14: Top: geoid height differences by degree for different solutions. Bottom: comparison between the non-robust IREML solution (dashed line) with the robust RVCE solution (solid line).

linear combination of the cofactor matrices should be more consistent to the residuals of the EGM96 matrix. We have investigated two eigenvalue decompositions. In the first decomposition, the VC-matrix up to degree $L = 75$ is split into four cofactor matrices; see the left panel of figure 7.15. The second decomposition first defines a cofactor matrix for the spherical harmonics with $l > 70$, and performs an eigenvalue decomposition on the remaining part $l \leq 70$. This is shown in the right panel of figure 7.15.

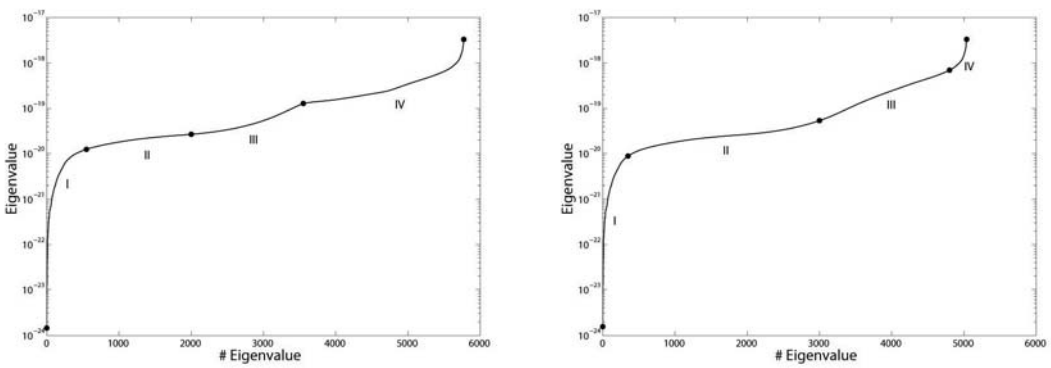


Fig. 7.15: Eigenvalues of the EGM96 variance-covariance matrix, with a (rather arbitrary) division into 4 groups. In the left figure, an eigenvalue decomposition is performed up to degree and order 75, in the right figure up to degree and order 70.

However, the re-weighting of these cofactor matrices did not improve the combination solution, compared to the RVCE solution, using one scale factor for the EGM96 VC-matrix. In terms of RMS geoid height differences with EIGEN-GL04C, the RVCE EVD solutions gave comparable, but slightly worse (7.0 cm) results than the RVCE solution without a EVD (6.9 cm), up to degree $L = 50$.

7.6.3 Augmentation of the functional model

In figure 7.14, we can see that some degrees of the CHAMP-only solution show large differences with the independent EIGEN-GL04C model. Now we will perform a statistical test for any inconsistencies between the CHAMP-only solution and the EGM96 global model. If we order the potential coefficients by degree, an inconsistency vector in degree l is augmented in the functional model as a vector of local (nuisance) parameters:

$$\begin{bmatrix} \underline{y}_{1,2:(l-1)} \\ \underline{y}_{1,l} \\ \underline{y}_{1,(l+1):L} \\ \hline \underline{y}_{2,2:(l-1)} \\ \underline{y}_{2,l} \\ \underline{y}_{2,(l+1):L} \end{bmatrix} = \left[\begin{array}{ccc|c} I_{2:(l-1)} & 0 & 0 & 0 \\ 0 & I_l & 0 & I_l \\ 0 & 0 & I_{(l+1):L} & 0 \\ \hline I_{2:(l-1)} & 0 & 0 & 0 \\ 0 & I_l & 0 & 0 \\ 0 & 0 & I_{(l+1):L} & 0 \end{array}\right] \begin{bmatrix} \underline{x}_{2:(l-1)} \\ \underline{x}_l \\ \underline{x}_{(l+1):L} \\ \hline \underline{\delta}_l \end{bmatrix} + \underline{e}, \qquad (7.9)$$

cf. section 3.2. We will test for such an inconsistency vector for every degree of the CHAMP-only solution. In the first test setup, we choose to test the CHAMP solution with the robustified stochastic model; in the second test setup, we test for inconsistencies in the same CHAMP solution, but in the non-robust combination with EGM96. The estimated test statistics, compared with the critical value (dependent on the number of local parameters), can be seen in figure 7.16. First we will look at the differences between

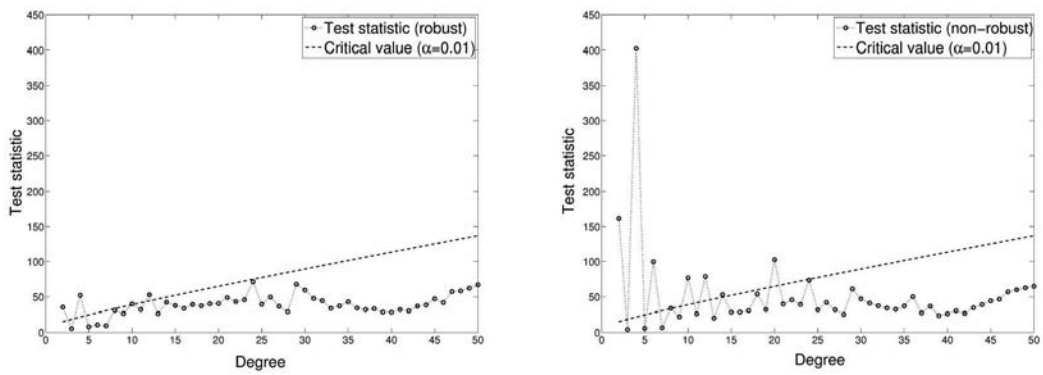


Fig. 7.16: Test statistics for the CHAMP solution for the robust (RVCE) combination (left), and the non-robust (VCE) combination (right).

the two plots. We can see that the high test statistics in the non-robust combination are much lower in the robust combination. As we would expect, the robust combination partly masks the inconsistencies between the two models. If one would like to test for inconsistencies, it is therefore better to look at the non-robust combination.

In this combination, we see significant inconsistencies in the even degrees up to $l = 14$. If we compare this to the geoid height differences by degree with EIGEN-GL04C (figure 7.14), we can conclude that there is a significant inconsistency in the degrees 2 and 4 of the CHAMP-only solution. The high test statistics for the even degrees between 6 and 14 are due to correlations with the degrees 2 and 4. These test statistics are therefore much lower in the robust combination test setup. Many groups made suggestions that there is something wrong with degrees 2 and 4 of the TUM kinematic orbits; see, e.g., Reubelt et al. [2004]. It was found to be an error in the reference frame during the processing of the CHAMP kinematic orbits. If we combine these results with figure 7.14, we can draw the conclusion that an inconsistency vector for degrees 2 and 4 should be estimated for the CHAMP solution in the combination with EGM96.

7.6.4 Stochastic model validation (second iteration)

The augmentation of the functional model not only improved the estimation of the coefficients of the degrees 2 and 4, or the coefficients, that are highly correlated to these coefficients but also improved the VCE and consequently the re-weighting of the data sets and the treatment of outlying coefficients (RVCE). The results in terms of geoid height differences with EIGEN-GL04C are shown in table 7.9.

Tab. 7.9: Statistics of the geoid height differences [m] between different combination solutions and EIGEN-GL04C.

Solution		$l \leq 10$		$l \leq 50$	
		Max.	RMS	Max.	RMS
Functional model not augmented	$\gamma_i = 1.0$	0.0312	0.0100	0.393	0.072
	IREML	0.0291	0.0095	0.480	0.075
	RVCE	0.0193	0.0056	0.417	0.069
Functional model augmented	$\gamma_i = 1.0$	0.0188	0.0060	0.348	0.063
	IREML	0.0189	0.0065	0.304	0.063
	RVCE	0.0195	0.0056	0.306	0.063

One can see that the robust solution partly absorbed the inconsistency in the lower degrees. After the implementation of the inconsistency vector, the results between the different solutions were comparable. The geoid height differences by degree between the RVCE solution, with and without an augmentation of the functional model, are shown in figure 7.17.

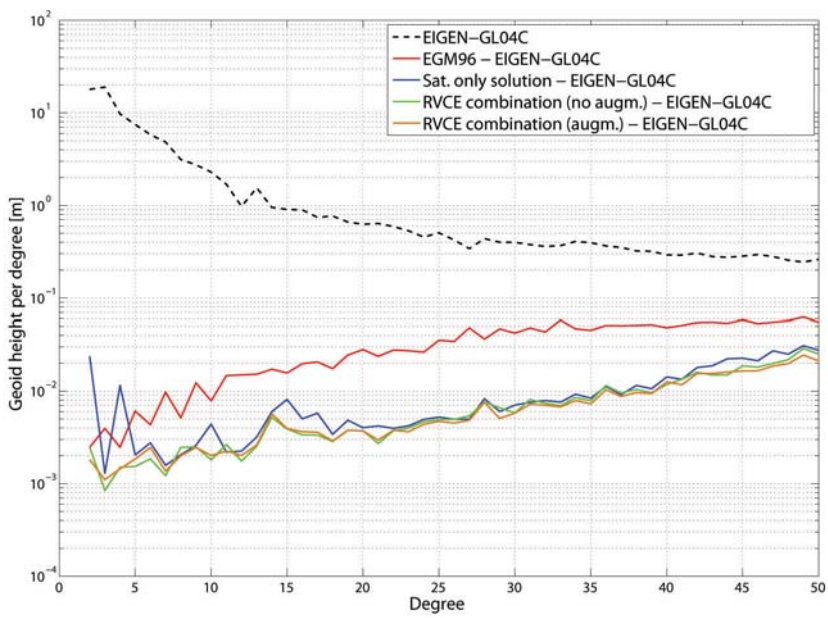


Fig. 7.17: Geoid height differences by degree [m] between different solutions and EIGEN-GL04C.

7.7 Summary and outlook

We have used CHAMP satellite data to test the proposed algorithms concerning the augmentation of the functional model, the validation of the stochastic model and the treatment of outliers. First, we computed CHAMP pseudo-observations using the energy balance approach. Monte Carlo variants of MINQUE, IREML and Lerch's subset solution method were compared to each other and to the Iterative Maximum Likelihood Estimator (IMLE). All four estimators converged to quite similar variance components, although Lerch's method gave less weights to some spurious data sets. However, the estimators were quite different in terms of computational costs involved. The MC-MINQUE and MC-Lerch methods were several orders slower than the fast MC-IREML method and the very fast IMLE method. As the IMLE method produces biased estimates, we chose to use the MC-IREML estimator to validate the stochastic model.

With a proper stochastic model, we could test for significance of the local parameters. All 6 parameters, i.e., a bias, a drift, and four parameters to account for periodic effect at 1 cpr and 2 cpr, were found to be significant. Correlation tests showed us that singularity problems could arise if the data sets are too short to distinguish between the different local parameters. This was however not the case in our test setup.

The VCE was used in a combined iterative setting, together with the outlier treatment techniques. Several techniques have been discussed and compared, including the conventional 3-sigma rule, M-estimation and the Cost Function Estimation (CFE), of which the latter was found to produce the smallest differences with the accurate (and independent) EIGEN-GL04C model. The Huber influence function was used in the estimation of the variance components.

A CHAMP-only model was computed using the full two year data set. Geoid height differences by degree (with EIGEN-GL04C) and tests for inconsistencies with EGM96 showed that the degrees 2 and 4 were badly determined. This was also found by other groups, who contributed this to an error in the reference frame during the processing of the kinematic orbits. We augmented the functional model in the combination of CHAMP with EGM96 to account for an inconsistency vector between the two data sets.

An attempt to divide the VC-matrix of EGM06 into several cofactor matrices using an eigenvalue decomposition and consequently rescale these cofactor matrices (with VCE) did not show better results than a simple rescaling of the entire EGM96 VC-matrix.

The robust combination of the CHAMP model and EGM96 could partly absorb the inconsistencies in the lower degrees. However, after the augmentation of the functional model to account for these inconsistencies, the robust combination showed quite similar results than the non-robust combination method.

Future work should include correlations between the CHAMP pseudo-observations, as these observations show correlations due to the derivation of the velocities from the kinematic positions. With a better stochastic model of CHAMP, we expect to obtain better estimates of the variance components of the EVD cofactor matrices of EGM96 and in this way improve the VC-matrix of EGM96.

8

Application 2: Joint inversion of global GPS time-series with GRACE gravity models

In this chapter we will combine weekly coordinate solutions of a global GPS network with monthly spherical harmonic GRACE solutions. Both 'observation' types are highly sensitive to temporal mass variations on the surface of the Earth, due to, e.g., the global water cycle.

The GRACE satellite mission (launched in 2002) provides us with monthly $L = 120$ solutions for the time-variable part of the Earth's gravity field. These time-varying gravity models can be directly inverted into surface mass changes. In this way, the GRACE mission has provided us with information on the seasonal hydrological signal in large river basins [Tapley et al., 2004a] and the decrease in ice mass in the polar regions [Velicogna and Wahr, 2006]. However, the low-degree spherical harmonic coefficients (especially $\Delta\bar{C}^{\sigma}_{20}$) of the monthly GRACE models are poorly resolved. This mainly disturbs the estimation in the equatorial regions and the polar regions, which are our regions of interest.

A combination with an independent data set, which can solve for these low-degree coefficients, is therefore necessary. Only recently [Blewitt et al., 2001], weekly time series of GPS site displacements were used to obtain estimates of the mass redistributions on the surface of the Earth. However, in case of globally supported basis functions, like spherical harmonics, this inversion is highly unstable, even for the very low degrees. A regularization or a combination with other data (e.g., GRACE) is therefore necessary. Monthly regularized GPS-only solutions (figure 8.13) show reasonable results, compared to the GRACE-only solutions; see figure 8.14. Moreover, weekly geocentre motions can be measured by GPS site displacements in a joint estimation with the coefficients of the surface mass changes.

Due to the different signal characteristics of GRACE and GPS, a joint inversion of both observation types [Kusche and Schrama, 2005] can improve the estimation of the low-degree spherical harmonics compared to the GRACE-only solutions and can stabilize the estimation of the geocentre motion and other local parameters needed in the estimation of the GPS inversion.

We will focus on the joint inversion of GPS and GRACE using VCE for the relative weighting and a robust estimation of the derived observations to detect and treat possible outliers. Other techniques to measure either the deformations of the Earth, such as InSAR [Galloway et al., 1998], or the change in sea surface heights from radar altimeter data and ocean bottom pressure (OBP) data [Wu et al., 2006], will not be considered.

8.1 The functional model of the joint inversion

The goal of the joint inversion is to estimate the coefficients $\Delta\bar{C}^{\sigma}_{lm}, \Delta\bar{S}^{\sigma}_{lm}$ (Eq. (2.32)) and consequently the surface mass redistributions, which can be directly translated to equivalent water heights; see also section 2.2. As the GRACE satellite gravity data is provided from degree 2, the common parameters in the joint inversion are the spherical harmonic coefficients from degree 2 to L. The functional model of the time-varying satellite gravity data (e.g., GRACE), therefore, reads

$$\underline{y}_1 = A_{1s}\underline{x}_s + \underline{e}_1 \tag{8.1}$$

with

$$\begin{aligned}
\underline{y}_1 &= \text{vector of observations: } [\ldots \Delta\bar{C}^{sat}_{lm}, \Delta\bar{S}^{sat}_{lm}, \ldots]^T \\
A_{1s} &= \text{diagonal design matrix, with diagonal elements } \frac{3\cdot\rho_w}{\rho_e}\frac{1+k'_l}{2l+1} \\
\underline{x}_s &= \text{vector of unknowns: } [\ldots \Delta\bar{C}^{\sigma}_{lm}, \Delta\bar{S}^{\sigma}_{lm}, \ldots]^T \\
\underline{e}_1 &= \text{vector of residuals,}
\end{aligned}$$

where $l = 2, \ldots, L$. Note that a reference model (mostly the average over a larger time span) needs to be removed from the GRACE coefficients.

The second observation group consists of all GPS stations with three observations $(\Delta h_i, \Delta e_i, \Delta n_i)$ for each station. A reference position needs to be subtracted from the GPS time series to obtain relative site displacements. To avoid inconsistencies with the GRACE observations, one should be careful to use the same time interval in the computation of the reference points as in the GRACE reference model.

The functional relationship of the height (Δh), eastward (Δe) and northward (Δn) displacements of station i reads respectively, cf. Eq. (2.41), Eq. (2.44) and Eq. (2.45):

$$\begin{aligned}
\Delta h(\vartheta_i, \lambda_i) = \; & R\frac{3\rho_w}{\rho_e}\sum_{l=1}^{L}\frac{h'_l}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda_i + \Delta\bar{S}^{\sigma}_{lm}\sin m\lambda_i]\bar{P}_{lm}(\cos\vartheta_i) \\
& + \underline{e}_{h_i}\cdot\underline{b} - R\cdot s,
\end{aligned} \tag{8.2}$$

$$\Delta e(\vartheta_i,\lambda_i) = \frac{R}{\sin\vartheta_i}\frac{3\rho_w}{\rho_e}\sum_{l=1}^{L}\frac{l'_l}{2l+1}\sum_{m=0}^{l} m\cdot[-\Delta\bar{C}^{\sigma}_{lm}\sin m\lambda_i + \Delta\bar{S}^{\sigma}_{lm}\cos m\lambda_i]\bar{P}_{lm}(\cos\vartheta_i)$$
$$+\underline{e}_{e_i}\cdot\underline{b} + \underline{e}_{n_i}\cdot\underline{\epsilon} \tag{8.3}$$

and

$$\Delta n(\vartheta_i,\lambda_i) = -R\frac{3\rho_w}{\rho_e}\sum_{l=1}^{L}\frac{l'_l}{2l+1}\sum_{m=0}^{l}[\Delta\bar{C}^{\sigma}_{lm}\cos m\lambda_i + \Delta\bar{S}^{\sigma}_{lm}\sin m\lambda_i](\frac{\partial}{\partial\vartheta}\bar{P}_{lm}(\cos\vartheta)|_{\vartheta=\vartheta_i})$$
$$+\underline{e}_{n_i}\cdot\underline{b} - \underline{e}_{e_i}\cdot\underline{\epsilon} \tag{8.4}$$

with

$\underline{e}_{h_i}, \underline{e}_{e_i}, \underline{e}_{n_i}$	= unit vectors in the local reference frame
$\underline{b}$	= Helmert transformation parameters
$\underline{\epsilon}$	= Helmert rotation parameters
s	= Helmert scaling parameter.

For each epoch, we estimate 7 Helmert parameters to account for residual datum inconsistencies in the weekly GPS solutions; see, e.g., Kusche and Schrama [2005] and Wu et al. [2006]. The degree-1 spherical harmonic coefficients $\Delta\bar{C}^{\sigma}_{10}$, $\Delta\bar{C}^{\sigma}_{11}$ and $\Delta\bar{S}^{\sigma}_{11}$ (= geocentre motion parameters) are not present in the functional model of the satellite gravity observations and should, therefore, also be considered as local parameters, together with the 7 Helmert parameters. We have chosen to work in the Centre of Figure (CF) frame and therefore set the degree 1 Load Love Numbers to $l'_l = 0.134$ and $h'_l = -0.269$ [Kusche et al., 2007]. More on the subject of reference frames can be found in Dong et al. [2003] and Blewitt [2003].

We will assume that there is a mass conservation within the surface layer σ_R, i.e. we define $\Delta\bar{C}^{\sigma}_{00} := 0$ as a constraint [Wahr et al., 1998]. This is a valid assumption if the atmospheric mass changes are considered to be included in the surface layer. These masses should, therefore, be present in both GPS and GRACE data.

8.2 Inversion of weekly GPS displacements

8.2.1 Description of the GPS data

First we will focus on weekly GPS-only solutions. The input data consists of 104 weeks (1260 to 1363) of GPS combination solutions for 158 sites, spanning the period of March 2004 to February 2006, as part of the network of the International GPS Service (IGS). Full covariance information is available and will be used in the solution. Temporal variations are obtained relative to a two year average (March 2004 - February 2006). More information on the IGS network solutions can be found in Ferland et al. [2000]. A height time series of a GPS station in the Amazon is shown in figure 8.1.

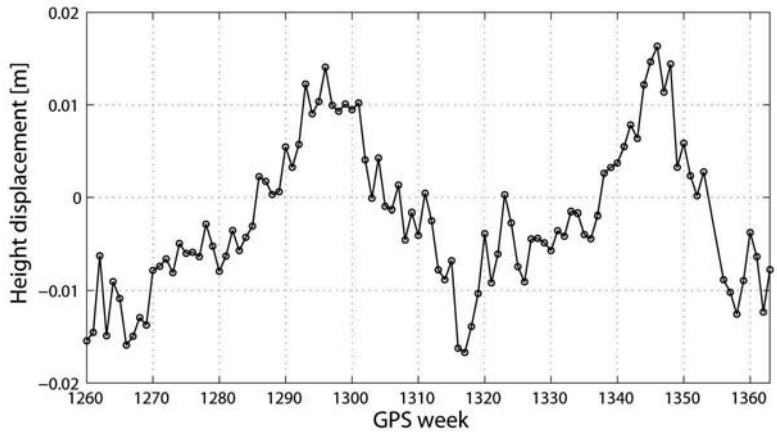


Fig. 8.1: Two-year (March 2004 - February 2006) time series of the height displacement (in meters) of the Amazon IGS GPS station in Brasilia, Brazil.

8.2.2 Constraining the solution

Due to the $(2l + 1)^{-1}$ term in Eqs. (8.1)-(8.4) and the behaviour of the Load Love Numbers for increasing degrees, the model of the surface mass redistributions becomes increasingly sensitive to errors in the high-degree coefficients in either the satellite gravity data or the GPS displacements [Blewitt and Clarke, 2003]. Moreover, the sparse and heterogeneous distribution of the GPS stations (figure 8.2) causes an unstable GPS-only solution, already from the lowest spherical harmonic degrees ($l > 3$). A solution to this is to use localizing basis functions which can be directly related to spherical harmonics. If one keeps to global representation using series of spherical harmonics, one needs to constraint the solution, which can be done by a low-degree truncation [Blewitt et al., 2001], by spatial averaging [Wahr et al., 1998; Kusche, 2007], or by regularization [Kusche and Schrama, 2005; Wu et al., 2006].

Low-degree truncation

We will combine point values (GPS measurements) with a continuous function, truncated to a certain degree L, and solve to this degree L. In this way, the high-degree ($l > L$) frequencies of the site displacements will enter into the truncated solution (*leakage error*). To reduce this error, we have to choose a sufficiently high degree L. However, the signal-to-noise ratio of the monthly GRACE solutions decreases for increasing L. As a first try, we have chosen to use a truncation degree $L = 14$ for the combination. In this way, we have prevented the bad resonance orders of the GRACE models (CSR, RL04) from entering into the solution.

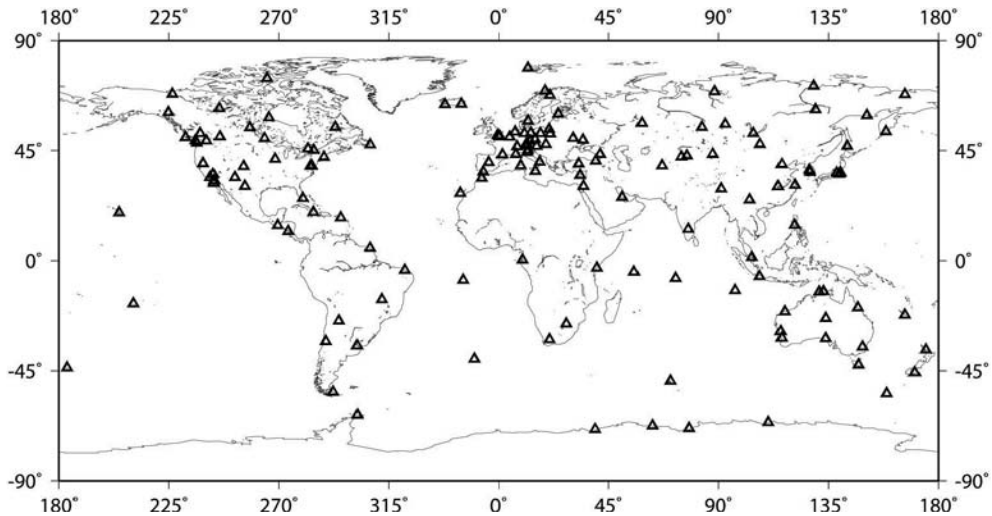


Fig. 8.2: GPS stations used in the inversion

Spatial averaging

High-degree errors can be smoothed in the space domain by taking a weighted average over a certain area. This can either be done by an isotropic weighting kernel [Wahr et al., 1998], or by a non-isotropic weighting kernel [Han et al., 2005cl; Cheng et al., 2006; and Kusche, 2007]. If the weighting kernel is isotropic, the averaging operation can be replaced by a weighting operation in the frequency domain, in which the weighting factors are only degree-dependent. The most frequently used isotropic weighting kernel is the Gaussian kernel, suggested by Jekeli [1981]. As we have truncated our solution to degree 14 and a spatial averaging would also smooth the signal itself, we have chosen not to perform any spatial averaging in this study.

Regularization

In Kusche and Schrama [2005], it was suggested to add a regularization matrix to the system of normal equations and in this way stabilize the solution. This regularization matrix is the result of a refined least-squares solution, in which we do not only minimize the vector of residuals $\underline{e} = \underline{y} - A\underline{x}$, but also the signal itself over the oceans. The cost function to be minimized reads:

$$\mathcal{J}_\alpha = \|\underline{y} - A\underline{x}\|^2_{Q_y^{-1}} + \alpha_o \int O(\vartheta,\lambda) \left(\frac{\Delta\sigma(\vartheta,\lambda)}{R\cdot\rho_w}\right)^2 \mathrm{d}\omega, \tag{8.5}$$

with $O(\vartheta,\lambda)$ the characteristic function over the oceans (1 over the oceans, 0 over land) and α_o the ocean damping parameter. The minimization of the signal over the oceans is valid as the oceanic mass variations are mainly due to diurnal and semi-diurnal ocean tides, which are corrected during the IGS GPS and GRACE data processing. The regu-

larization matrix can be computed as [Kusche and Schrama, 2005]:

$$[\Xi_{\mathcal{O}}]_{lm,l'm'} = \int_{\sigma} O(\vartheta,\lambda)\bar{Y}_{lm}(\vartheta,\lambda)\bar{Y}_{l'm'}(\vartheta,\lambda)\mathrm{d}\omega, \tag{8.6}$$

with $\bar{Y}_{lm}(\vartheta,\lambda)$ defined in appendix A. The system of normal equations then reads

$$\left[A^T Q_y^{-1} A + \alpha_o \cdot \Xi_{\mathcal{O}}\right] \hat{\underline{x}}_\alpha = A^T Q_y^{-1} \underline{y}. \tag{8.7}$$

We can see this as a kind of Tikhonov regularization, using pseudo-observations (valued zero) with the matrix $\Xi_{\mathcal{O}}^{-1}$ as its variance-covariance matrix. The propagation of this VC matrix on a spatial grid by computing $\sqrt{\underline{b}_P^T \Xi_{\mathcal{O}}^{-1} \underline{b}_P}$, with $\underline{b}_P = [\ldots \bar{Y}_{lm} \ldots]^T$, is shown in figure 8.3.

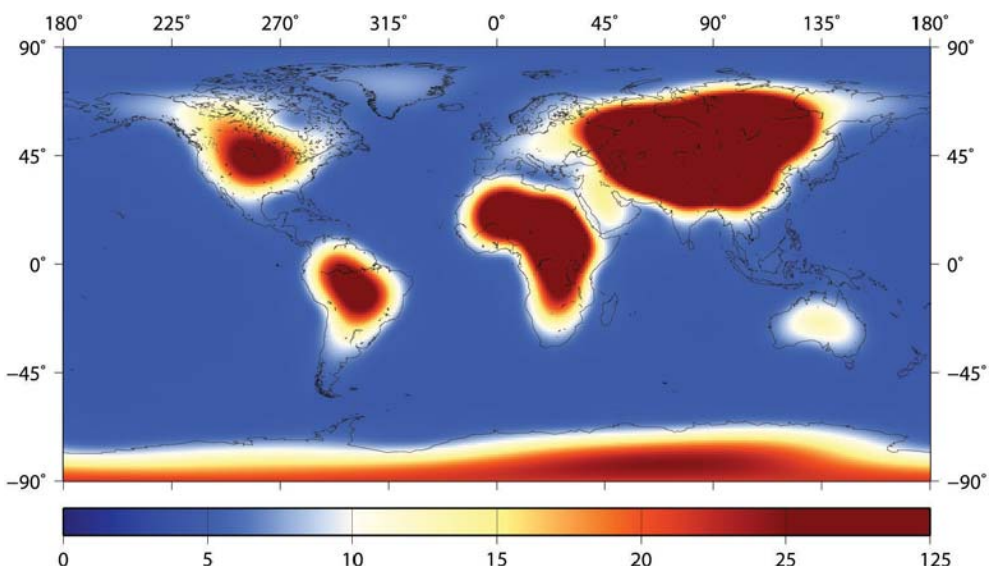


Fig. 8.3: Error propagation of $\Xi_{\mathcal{O}}^{-1}$ onto a spatial grid, $L = 14$.

In Wu et al. [2006] a combined global model of atmosphere, oceans and land hydrology is used as prior information to regularize the solution.

8.2.3 Weekly solutions

We have computed regularized weekly solutions up to degree 14 for all 104 GPS weeks. The ocean damping parameter α_o was estimated using IREML and had values of the order of 10^4, see figure 8.4. In Kusche and Schrama [2005], this parameter was obtained by looking at the spatial variability of the signal $\Delta\sigma$ over the oceans. They found, for a solution truncated to $L = 7$, the value of $\alpha_o = 0.9 \cdot 10^5$ from estimates of an ECCO ocean bottom pressure model and $\alpha_o = 0.3 \cdot 10^5$ from GRACE monthly solutions.

The regularized weekly solution of week 1299 is shown in figure 8.5. We can compare this weekly solution with the monthly GRACE solution (December 2004), see

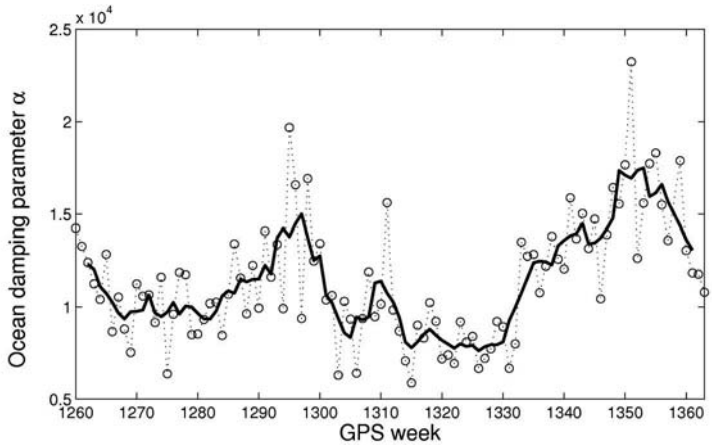


Fig. 8.4: Ocean damping parameter α_o for different GPS weeks, estimated from IREML for weekly GPS-only solutions, without the treatment of possible outliers. The solid line represents the moving average over 5 consecutive weeks.

figure 8.14. The main features of the GRACE solution are already visible in the regularized weekly solution of week 1299, such as the large mass variations in the Amazon River basin. Large inconsistencies with respect to GRACE occur in continental areas with a sparse distribution, e.g., Africa. More on this topic can be found in section 8.2.7. We will now try to reduce the effect of possible outliers on the solution.

8.2.4 Outlier treatment

We will compare the conventional 3-sigma rule with the robust M-estimation (Huber, $k = 1.7$) and apply IREML to rescale the VC-matrix per week. The estimation of the stochastic model and the treatment of the outliers are performed in an unregularized test setup. This is an iterative process, which converges within a few iterations. The ocean function (with the estimated parameter from the solution without any outlier treatment as fixed ocean damping parameter) is then used to regularize the solution. The difference of the solution, using the conventional 3-sigma rule, with the solution, using no outlier treatment, can be found in figure 8.6.

As the inversion is unstable in areas with only a few GPS sites, these areas will destabilize even further if outlying GPS sites are removed. Such a destabilization will deteriorate all spherical harmonic coefficients and therefore also have influence on remote areas. In figure 8.6 it is clearly visible that the removal of some observations in Europe and Asia has an effect on the surface mass estimates in South America and Africa. It is therefore better to use the robust M-estimation technique to down-weight outlying observations instead of removing them. The influence of this down-weighting on the solution can be found in figure 8.7.

If we compare figure 8.7 with figure 8.6, we can see that a robust treatment of the outlying observations mainly affects the areas in the proximity of these observations.

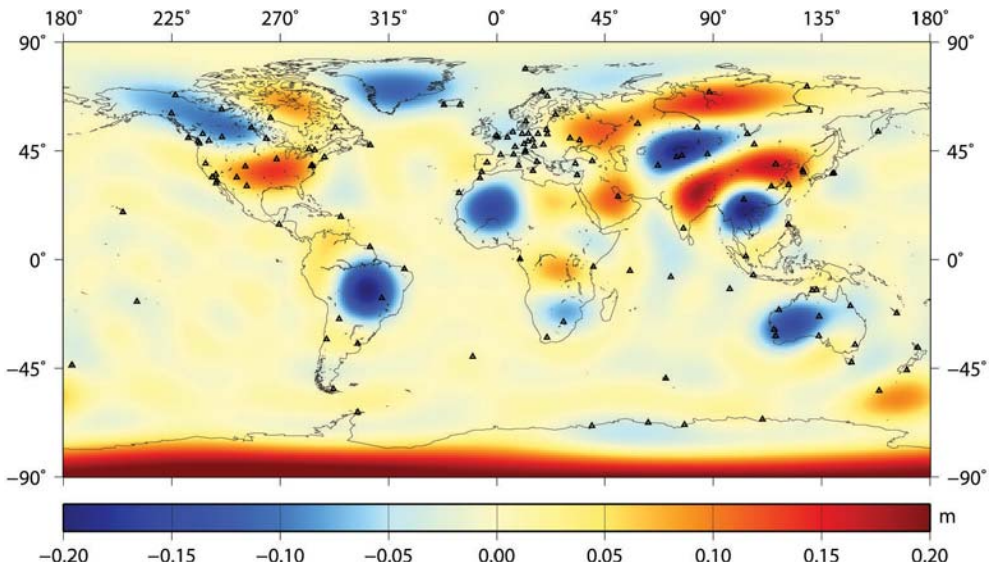


Fig. 8.5: Weekly regularized GPS-only solution for week 1299, expressed in terms of equivalent water heights [m], truncated at degree 14. No outlier treatment has been performed.

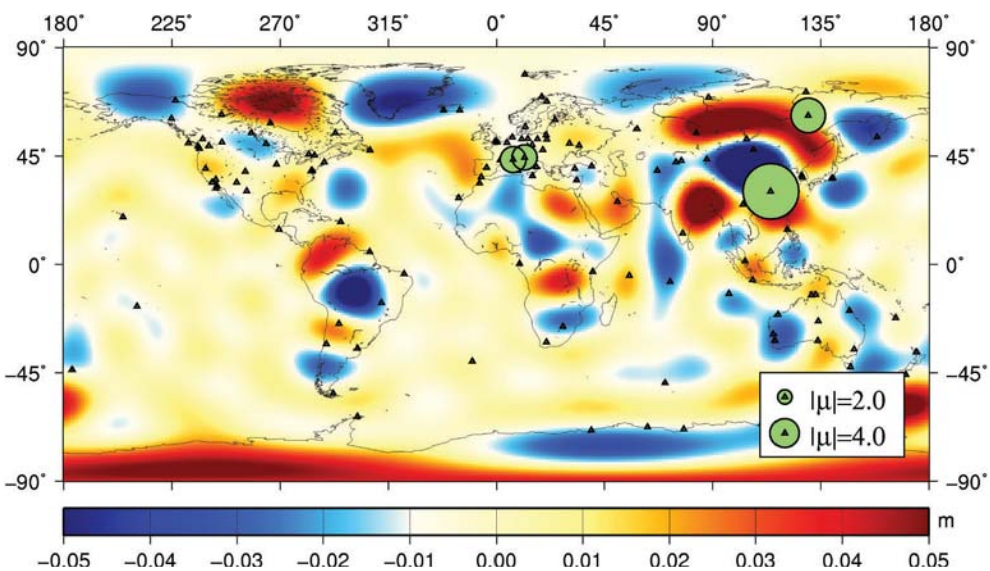


Fig. 8.6: Difference of the solution, using the 3-sigma rule, with the solution, using no outlier treatment (figure 8.5). The green circles display those observations, which are removed, with the size of the circles indicating the value of the test statistic.

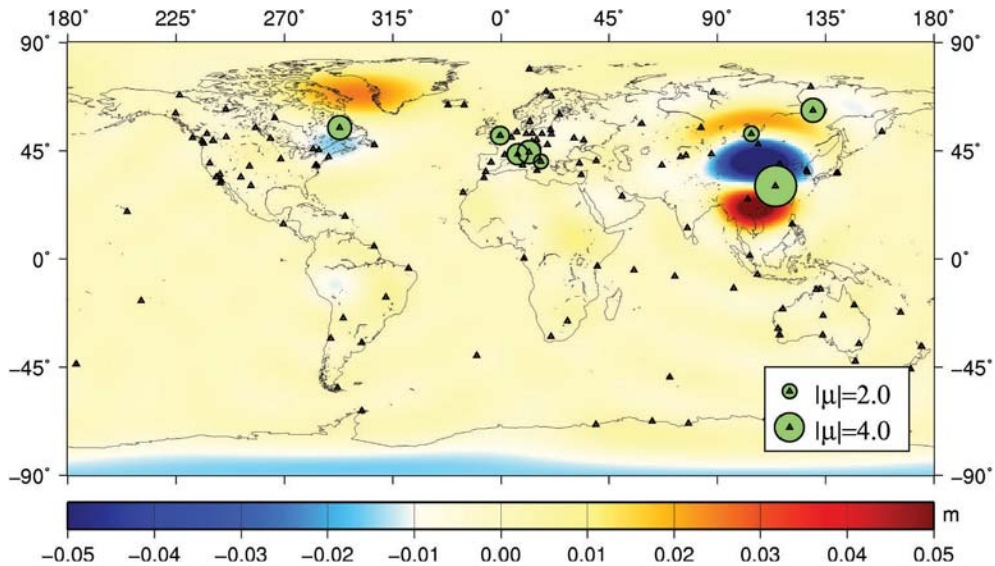


Fig. 8.7: Difference of the solution, using the robust M-estimation technique ($k = 1.7$) with the solution using no outlier treatment (figure 8.5). The green circles display the observations that are down-weighted, with the size of the circles indicating the value of the test statistic.

We therefore choose to use the robust M-estimation technique as the outlier treatment technique in this study.

8.2.5 Covariance information

We will now discuss how the covariances in the stochastic model of the GPS measurements contribute to the solution. The omission of these covariances (by setting them to zero) will disturb the estimation of the vector of unknowns, as the system of normal equations will change. Moreover, it will disturb the estimation of the stochastic model by VCE and consequently the treatment of the outliers using either the conventional 3-sigma rule or the robust M-estimation.

In figure 8.8 the robust solution is given for GPS week 1299, neglecting the covariances in the stochastic model. In the bottom panel one can see the difference with the robust solution using full variance-covariance information.

As we would expect, the continental areas where no GPS stations are located show the largest differences. The differences can reach up to 10 cm, which is half of the amplitude of the temporal signal. The covariances connect the information in the GPS stations and therefore play an important role in the estimation of the signal between these GPS stations. Neglecting the covariances mainly disturbs these regions.

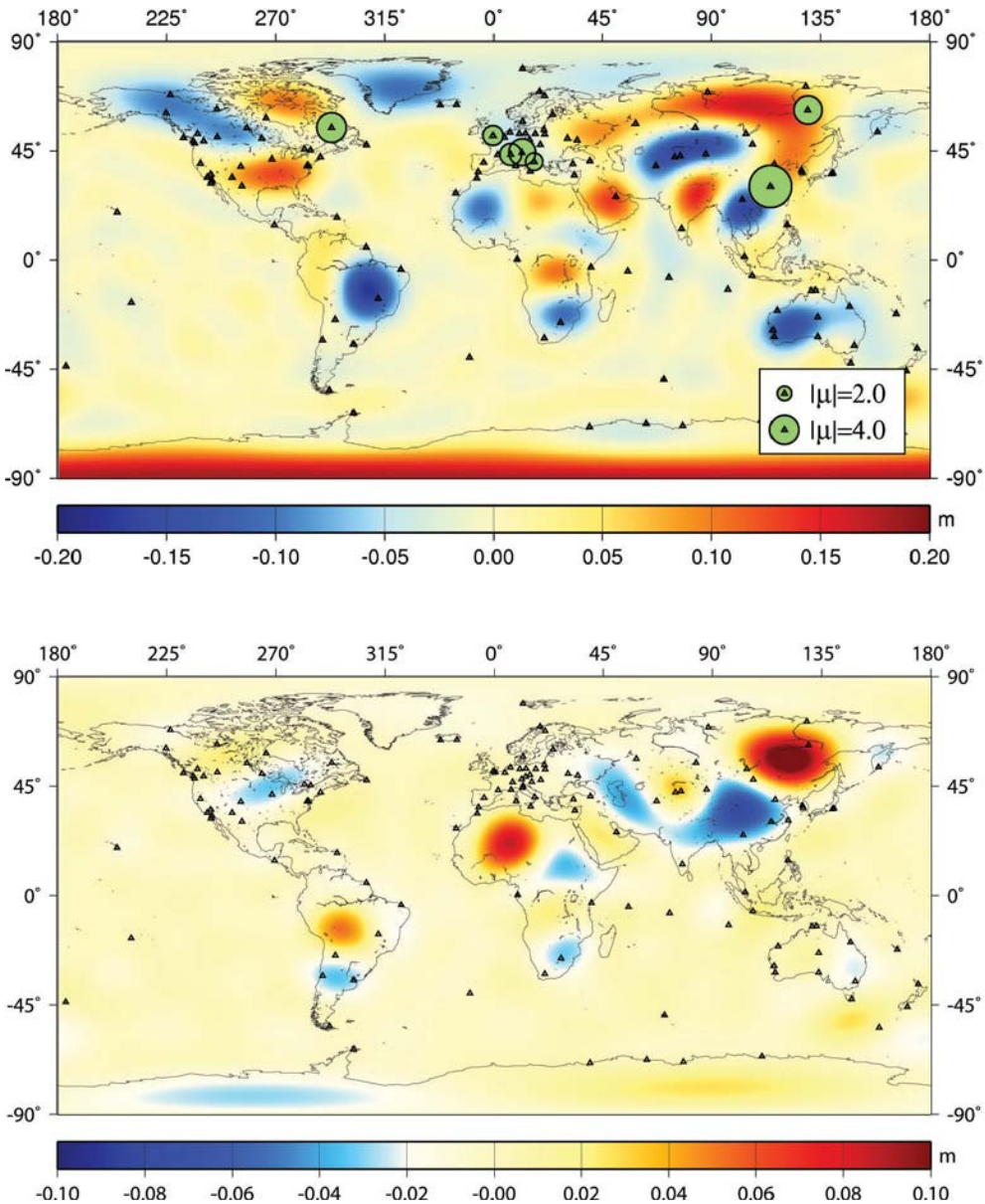


Fig. 8.8: The robust M-estimation GPS-only solution for week 1299, using a diagonal variance-covariance matrix for the GPS observations (top), and the difference of this solution with the robust solution, using full variance-covariance information (bottom). Note the different scale bars

8.2.6 Monthly solutions

We have used IREML to weight the different (4 or 5) weekly solutions in order to compute monthly GPS solutions. The ocean damping parameters were again estimated by IREML and were of the order of 10^3. These under-estimated ocean damping parameters caused the solution to oscillate over the oceans. We have therefore chosen to constrain the ocean damping parameter to $0.4 \cdot 10^5$, based on weekly solutions (figure 8.4), ECCO estimates and GRACE estimates [Kusche and Schrama, 2005].

A possible reason for these under-estimated damping parameters is the increased influence of the GPS observations. These GPS stations contain 4 to 5 times more observations, decreasing the trace of the group influence matrix of the ocean damping function. The solution will therefore be closer to the GPS observations, decreasing the weight of the damping function.

We have estimated monthly GPS-only unregularized solutions using 4 or 5 weeks per month. The robust M-estimation method (Huber, $k = 1.7$) is used to down-weight possible outliers. At convergence, the solution is regularized by the ocean damping function with a fixed ocean damping parameter of $0.4 \cdot 10^5$. The results are shown in figure 8.9.

8.2.7 Modified regularization matrix

We have shown that estimates are obtained of the equivalent water heights up to degree and order 14 from only GPS site displacements as input data. However, areas with only a few GPS stations show large oscillations, which can not be attributed to the signal itself. These areas, e.g., Northern Africa, should be damped by regularization, as was done over the oceans.

We have therefore modified the regularization matrix to damp the signal in those areas where only a small signal is expected. We have used two years (same time span) of monthly GRACE models (CSR, Release 04) to estimate the variability of the signal. More on these GRACE models can be found in the section 8.3. The 24 monthly grids ($L = 14$) are compared to each other, and for each grid point a root mean square value is computed. The results are shown in figure 8.10.

We will now have to choose a threshold, which defines the areas where signal can be expected and the areas in which the signal should be damped by regularization. A threshold of 3.5 cm seems the best compromise between taking as much signal into account over the continents and damping the oceans as much as possible. In this way, the ocean function $O(\vartheta, \lambda)$ is replaced by a signal constraint function $S(\vartheta, \lambda)$. With the new boundaries of $S(\vartheta, \lambda)$, a modified regularization matrix Ξ_S ($L = 14$) is computed. The propagation of its inverse Ξ_S^{-1}, i.e., the VC-matrix of the pseudo-observations (values zero), onto a global grid is shown in figure 8.11.

This modified regularization matrix will stabilize some unstable regions, such as Northern Africa, which is a continental region but with only a few GPS stations. The improvement of the regularized solution in this area can be found in figure 8.12.

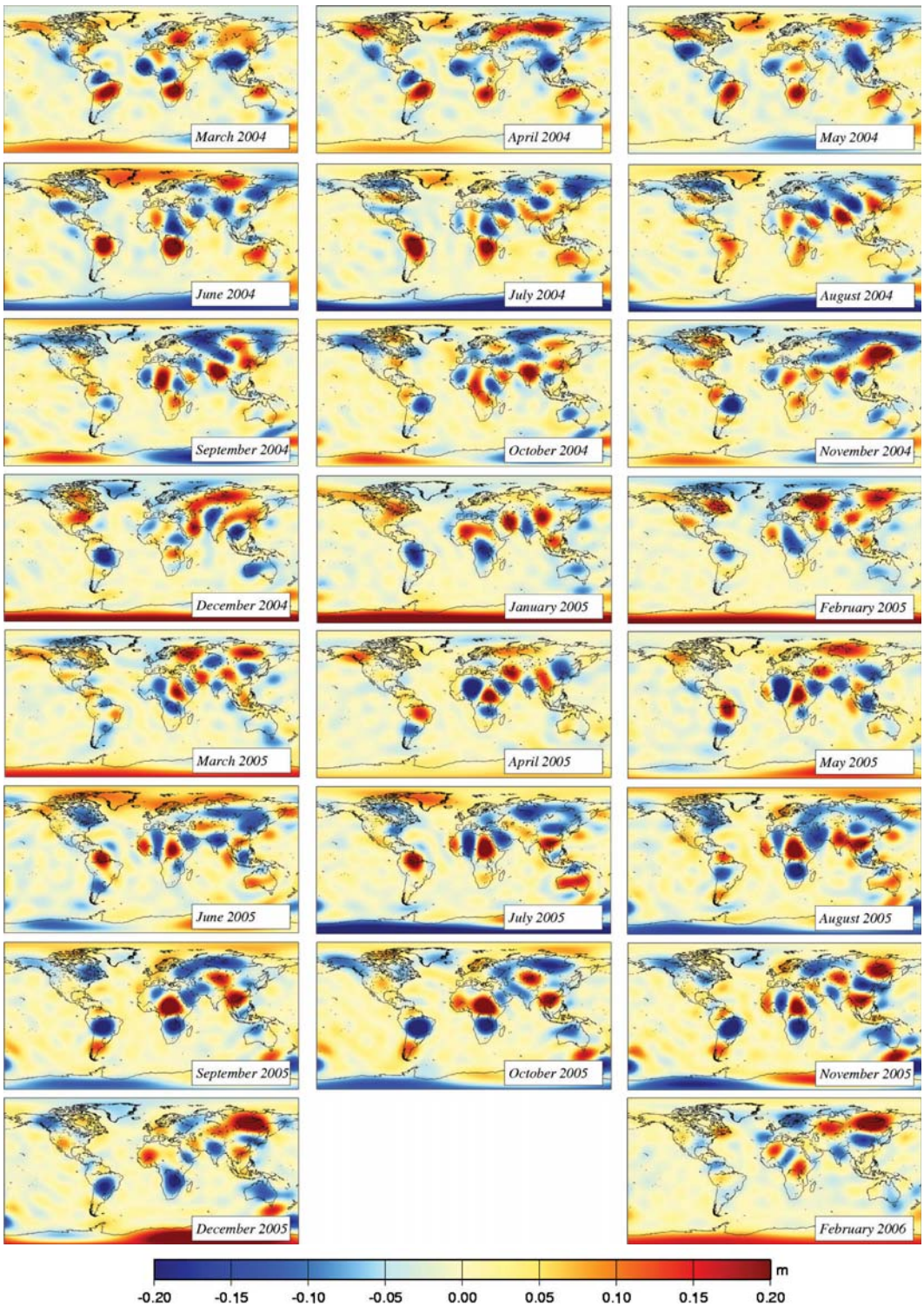


Fig. 8.9: Equivalent water heights [m] from monthly regularized GPS-only solutions (ocean damping) using robust M-estimation (Huber, $k = 1.7$) complete to degree $L = 14$.

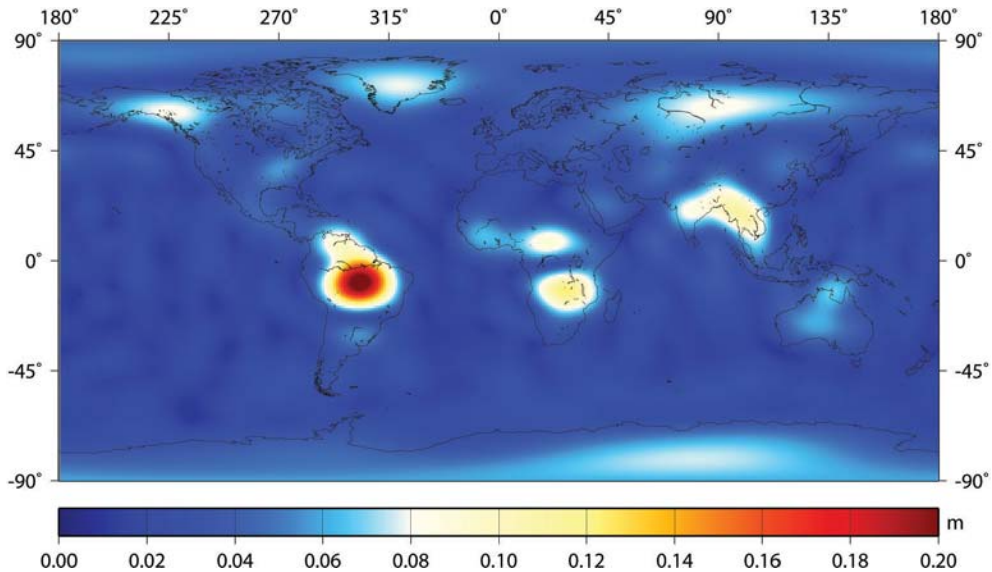


Fig. 8.10: Root mean square of the variations in equivalent water heights, retrieved from two years of monthly GRACE models (CSR, Release 04) up to degree and order 14.

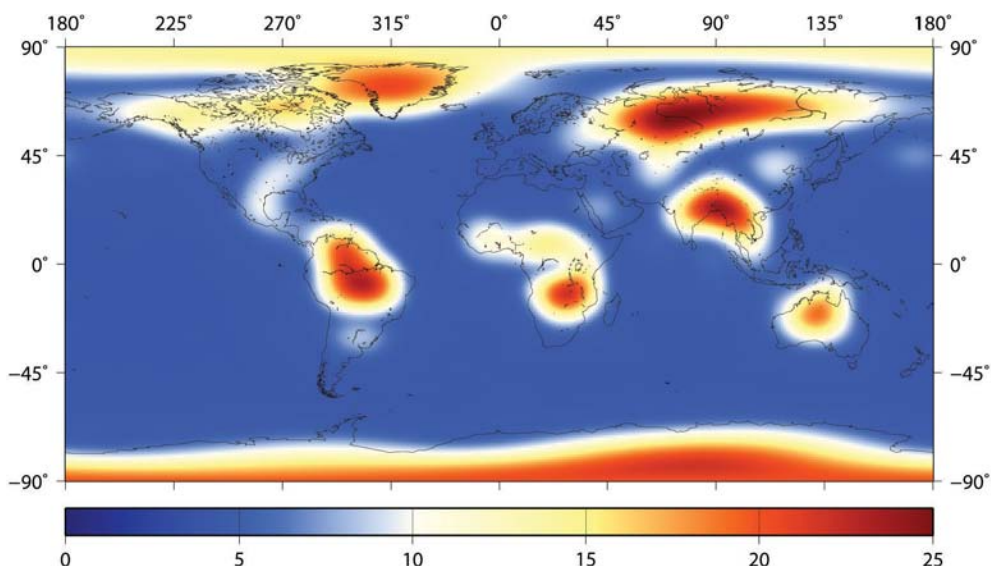


Fig. 8.11: Error propagation of $\Xi_{\mathcal{S}}^{-1}$ onto a spatial grid, $L = 14$.

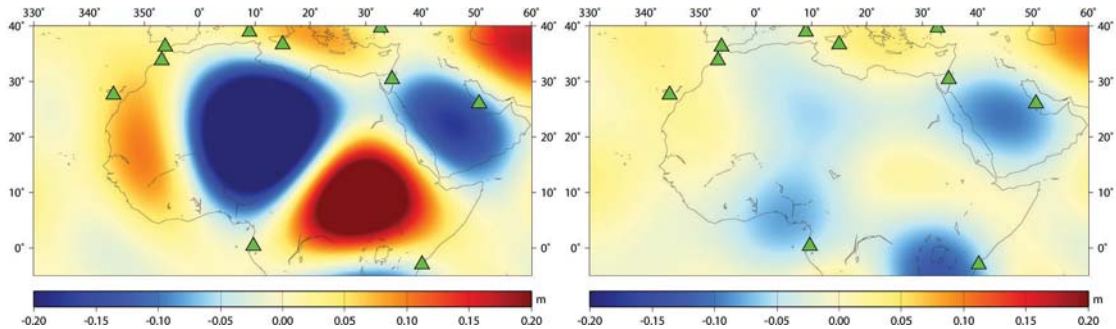


Fig. 8.12: Weekly GPS-only solution (week 1319) over Northern Africa, in terms of equivalent water heights [m], truncated to degree 14. Left: regularization with the ocean damping matrix $\Xi_{\mathcal{O}}$. Right: regularization with the signal constraint matrix $\Xi_{\mathcal{S}}$.

The robust monthly GPS solutions, using this new regularization matrix, are shown in figure 8.13. These solutions agree much better with the GRACE solutions (figure 8.14) than the ocean damped GPS solutions.

8.3 Inversion of monthly GRACE gravity models with weekly GPS displacements

One can see the combination of monthly GRACE solutions with weekly GPS site displacements as a regularization of the GPS-only solution by GRACE, making the previous regularization methods obsolete, or one could see the GPS data set as extra information on the low-degree spherical harmonics, which are poorly solved by GRACE. In this section, we will perform such a combination, using VCE methods to re-weight the different data sets and robust methods to treat possible outliers.

8.3.1 Description of the GRACE data

From the web-site of the Center of Space Research (CSR) [CSR, 2008], we obtained monthly estimates of the gravity field of the Earth (Release 04 GSM files). As non-tidal contributions of the oceans and the atmosphere (GAC files) are subtracted from the monthly gravity solutions, we need to add these back to the monthly models by the addition of the GAC files to the GSM files, as this signal is not removed from the GPS solutions. The average over the 24 GSM+GAC solutions (time span March 2004 - February 2006) is subtracted from the gravity solutions. In this way, the obtained GRACE gravity models should be consistent with the 104 weekly GPS IGS solutions. The GRACE CSR (Release 04) models, complete to degree 14, can be found in figure 8.14.

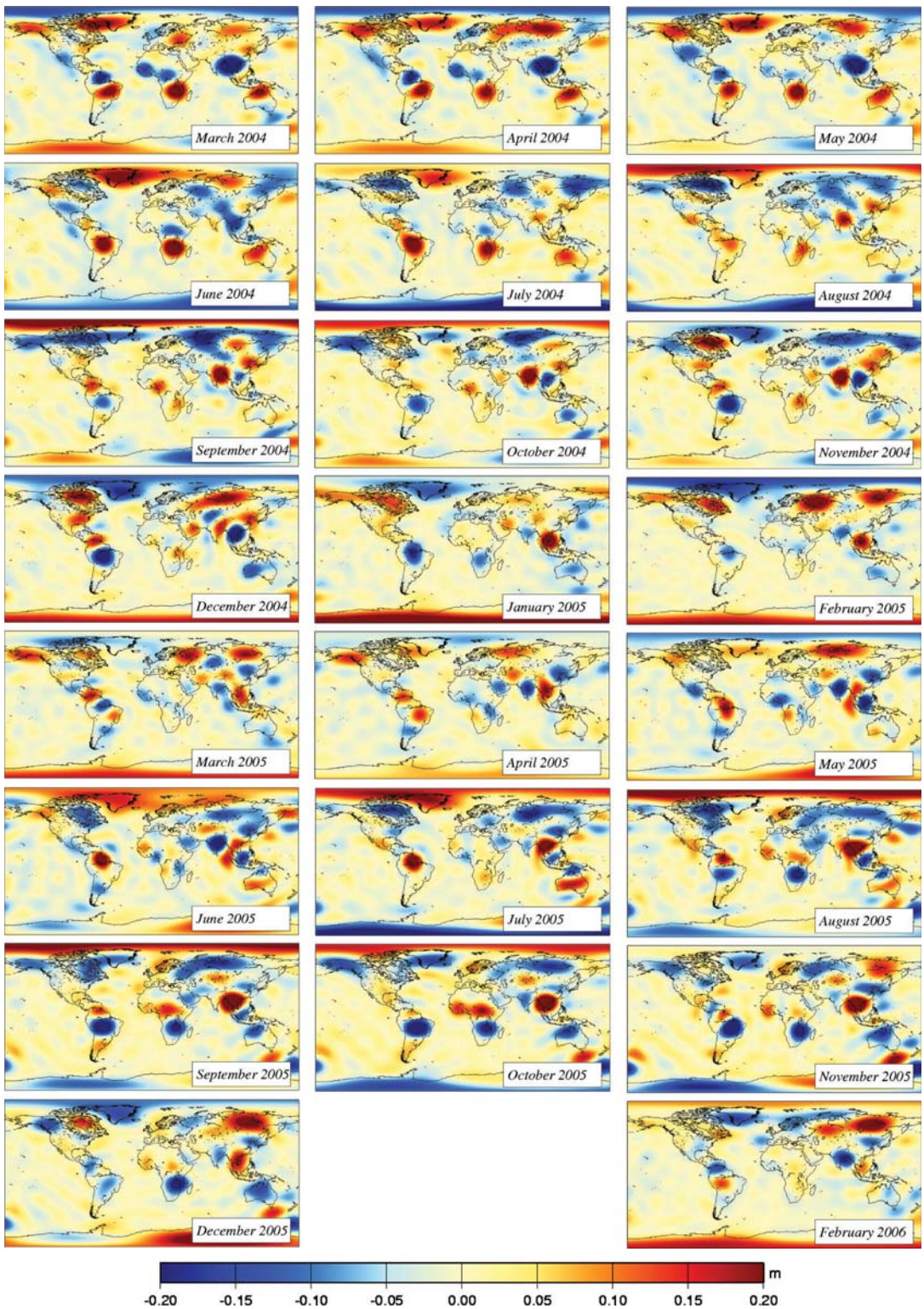


Fig. 8.13: Equivalent water heights [m] from monthly regularized GPS-only solutions (signal constraint) using robust M-estimation (Huber, $k = 1.7$), $L = 14$.

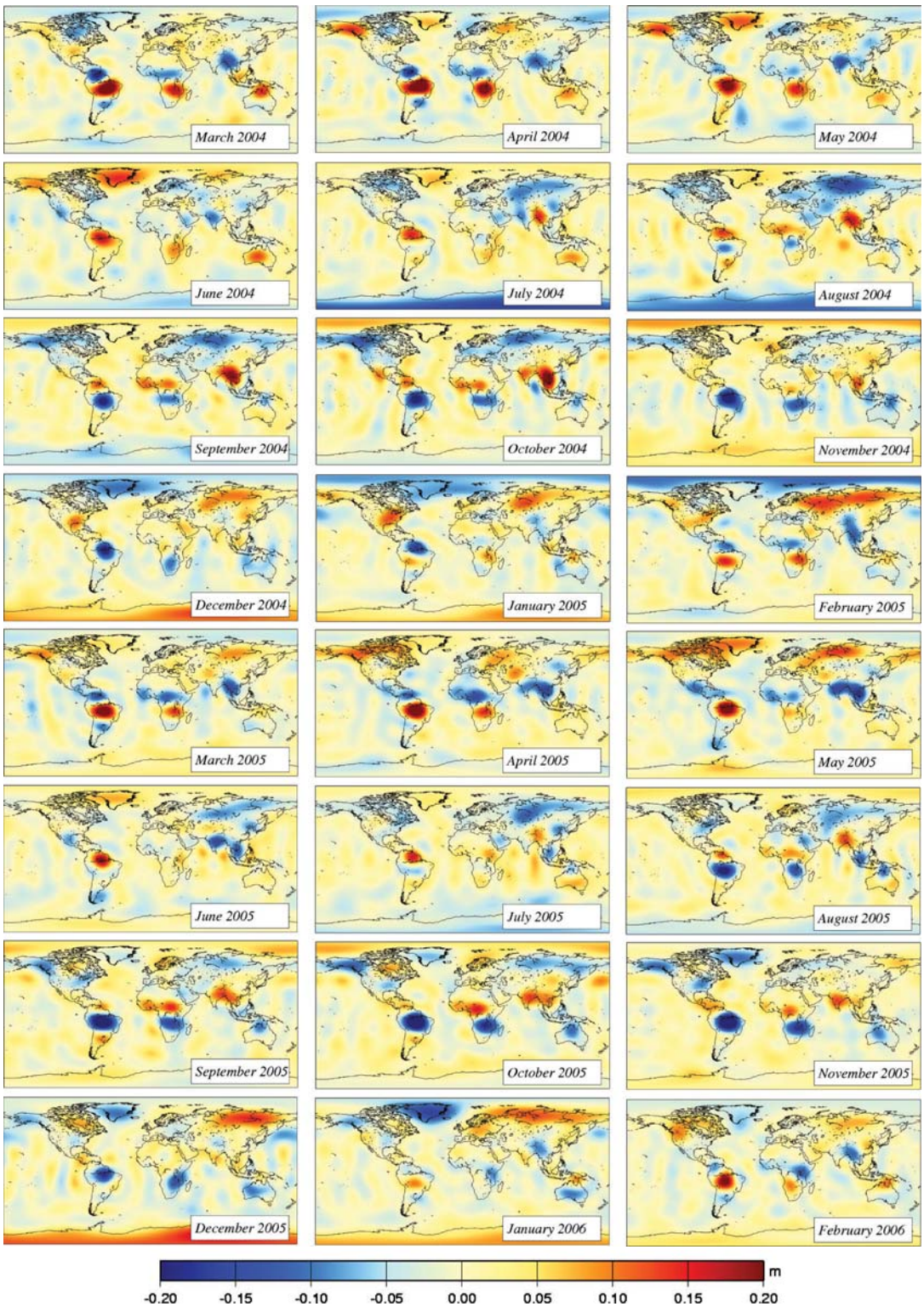


Fig. 8.14: Equivalent water heights computed from the GRACE monthly solutions (CSR, RL04), truncated to degree and order $L = 14$.

8.3.2 GPS/GRACE combination

As only the variance information and not the covariance information is publicly available for the CSR RL04 global gravity models, we have chosen to use the full variance-covariance matrix of September 2006 (which is publicly available) as the cofactor matrix for every month. This is a valid assumption as the structure of the variance-covariance matrix is expected to be quite homogeneous for all 24 months. A rescaling is done using VCE.

We will use GRACE and GPS in a joint inversion of monthly models of surface mass variations. In this way, GRACE will regularize the GPS normal matrices, making an ocean damping regularization or signal constraint regularization obsolete. For all GPS observations and GRACE spherical harmonic coefficients, robust weights are estimated in a combination with IREML to rescale the corresponding cofactor matrices.

However, due to inconsistencies between the two data sets in the lower degrees (figure 8.15), the GRACE solution gets a low weight in the first iteration. The accumulated normal matrix gets destabilized, producing high values over the oceans. In the next iteration, the weight of the GRACE solution is even further decreased. At convergence, the GRACE weight is decreased to zero with an unregularized monthly GPS solution as a result. Generalized cross-validation could not solve this problem.

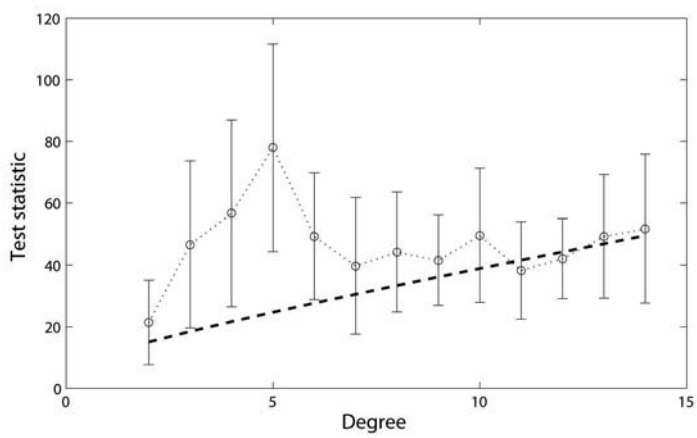


Fig. 8.15: Mean values (over two years) for the test statistics, testing for an inconsistency vector by degree between the GRACE monthly solutions and the monthly (robust) signal constraint regularized GPS solutions, with the error bars indicating the standard deviation over the two years. The dashed line represents the critical value ($\alpha = 0.01$).

We have therefore chosen to use the signal constraint regularization matrix to obtain proper relative weights for the GRACE model and the weekly GPS data sets. The relative weights of GPS and GRACE were then used to compute a GPS/GRACE combined solution (without any signal constraint).

The inclusion of GPS data mainly improves the low-degree spherical harmonic coefficients. The difference between the combined GPS/GRACE solution and the GRACE solution for October 2005 is shown in figure 8.16.

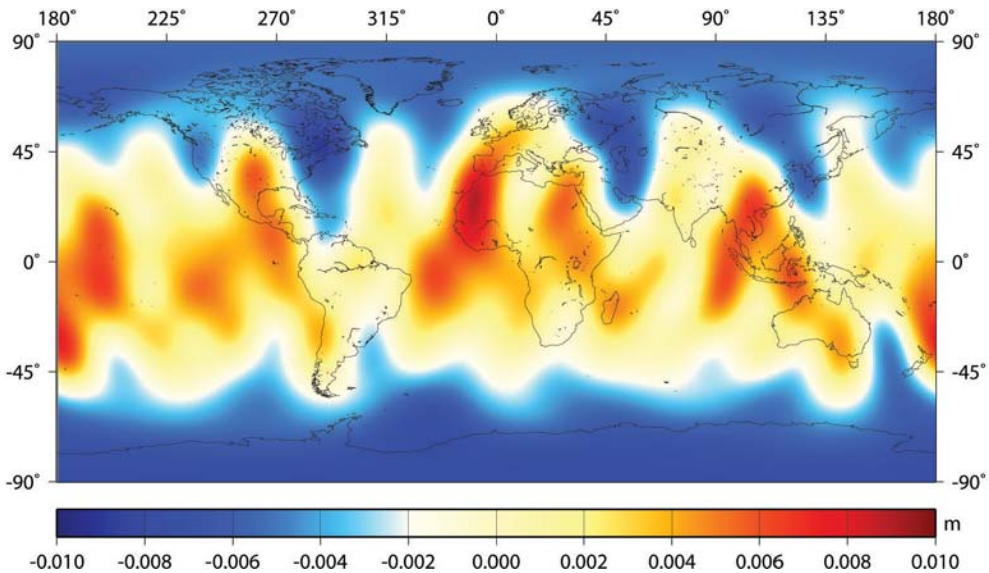


Fig. 8.16: Difference, in terms of equivalent water heights, between the combined GPS/GRACE solution and the GRACE solution for October 2005, truncated to degree 14.

Some months, like June 2005, show differences up to several centimetres. The improvements mainly come from the very low-degree spherical harmonic coefficients, especially from $\Delta\bar{C}^{\sigma}_{20}$. This can be clearly seen in figure 8.17, in which we have plotted the diagonal elements of the accumulated group influence matrix of GPS. This is a measure of the contribution of the data set on the estimation of the coefficients. The influence of GPS on the estimation of $\Delta\bar{C}^{\sigma}_{20}$ was (on average) 0.88. The influence numbers are in close agreement with studies by Kusche and Schrama [2005] and Jansen et al. [2008a].

The difference, in terms of equivalent water heights, between the combined GPS / GRACE solution and the GPS-only solution for October 2005 is shown in figure 8.18. The largest differences can be found in those areas with high signal amplitudes.

8.4 Evaluation of the local parameters

In addition to the spherical harmonic coefficients $\Delta\bar{C}^{\sigma}_{lm}$ and $\Delta\bar{S}^{\sigma}_{lm}$, we need to estimate 10 extra (local) parameters per week. Three of these parameters are connected to the degree 1 spherical harmonic coefficients and represent the geocentre motion, which can be estimated by GPS and is fixed for the monthly GRACE models. The geocentre

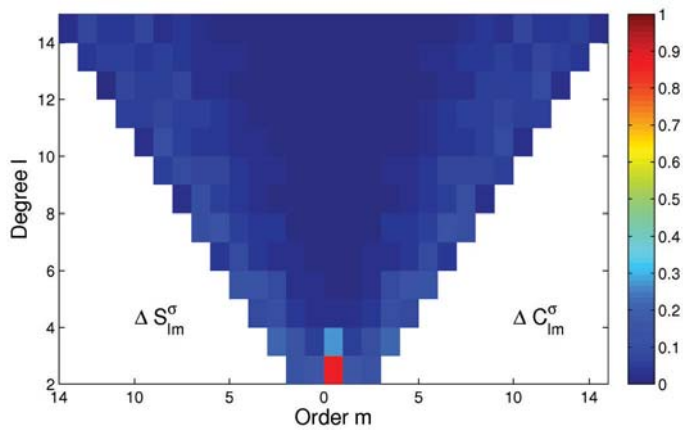


Fig. 8.17: Diagonal elements of the accumulated group influence matrix of GPS, averaged over all months, for each spherical harmonic coefficient.

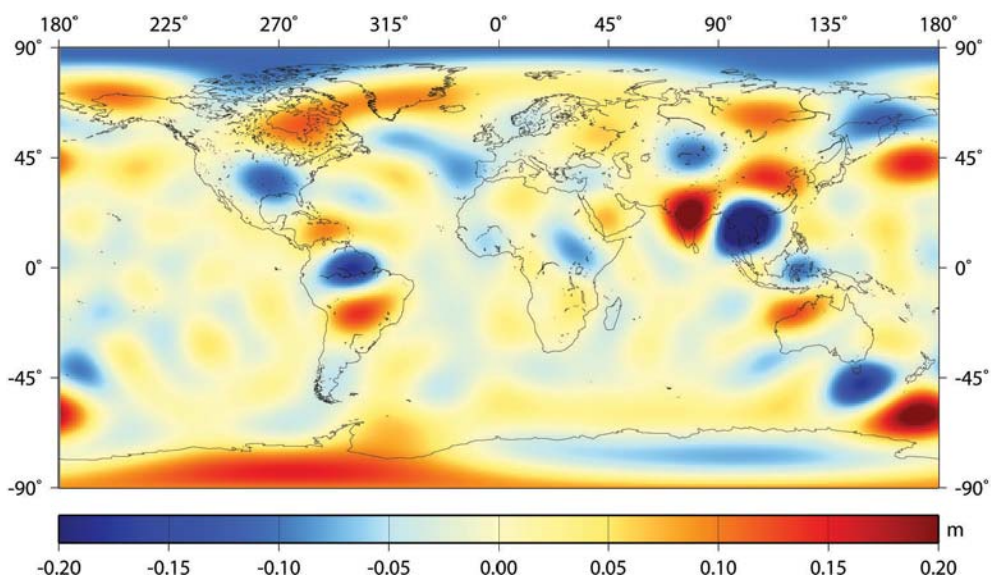


Fig. 8.18: Difference, in terms of equivalent water heights, between the combined GPS/GRACE solution and the GPS-only solution for October 2005, truncated to degree 14.

motions X, Y and Z can be retrieved by

$$\begin{aligned} X &= \frac{1}{M}\frac{4\pi a^4 \rho_w}{3}\Delta\bar{C}_{11}^{\sigma} = a\frac{\rho_w}{\rho_e}\Delta\bar{C}_{11}^{\sigma} \\ Y &= \frac{1}{M}\frac{4\pi a^4 \rho_w}{3}\Delta\bar{S}_{11}^{\sigma} = a\frac{\rho_w}{\rho_e}\Delta\bar{S}_{11}^{\sigma} \\ Z &= \frac{1}{M}\frac{4\pi a^4 \rho_w}{3}\Delta\bar{C}_{10}^{\sigma} = a\frac{\rho_w}{\rho_e}\Delta\bar{C}_{10}^{\sigma} \end{aligned} \tag{8.8}$$

The other 7 parameters account for a Helmert transformation of the GPS frame with respect to the GRACE reference frame.

8.4.1 Test for significance

If we want to draw any conclusions from the estimates of the local parameters, we first need to test if these parameters are significant; see section 3.2.2. The weekly test statistics of the local parameters are shown in figure 8.19.

Although the majority of the test statistics are below the critical value, partly due to the oscillations of the signal around zero, the high test statistics justify the estimation of these local parameters. If we would not estimate parameters for the residual Helmert transformation, they will alias into the estimation of the geocentre motion.

8.4.2 Motion of the geocentre

The motion of the geocentre can either be estimated in the regularized GPS-only estimation or the combined GPS/GRACE estimation. Figure 8.20 shows that for the majority of the weeks both estimates of the geocentre are comparable to each other.

However, for some weeks $(1339 - 1355)$ that correspond to the period September 2005 to December 2005, both Y and Z components show an offset between the GPS-only estimates and the GPS/GRACE estimates. If we look at the test statistics for this period, we can see that the geocentre motions are not significant for this period and that the ϵ_1 and ϵ_2 Helmert rotation parameters show large values for the test of significance. Therefore, a possible reason for the offset is that the parameters of the geocentre motion are poorly resolved in the combined GPS/GRACE estimation.

The Helmert transformation parameters are small ($\sigma \approx 5.0 \cdot 10^{-4}$) and do not show seasonal cycles. Therefore, it is not expected that these parameters partly absorb the motion of the geocentre [Kusche and Schrama, 2005].

8.5 Summary and outlook

The global water cycle generates a time-varying gravity signal due to land hydrology near the Earth's surface and mass changes in the atmosphere and cryosphere. The GRACE mission is highly sensitive to this time-varying signal and its data is used to

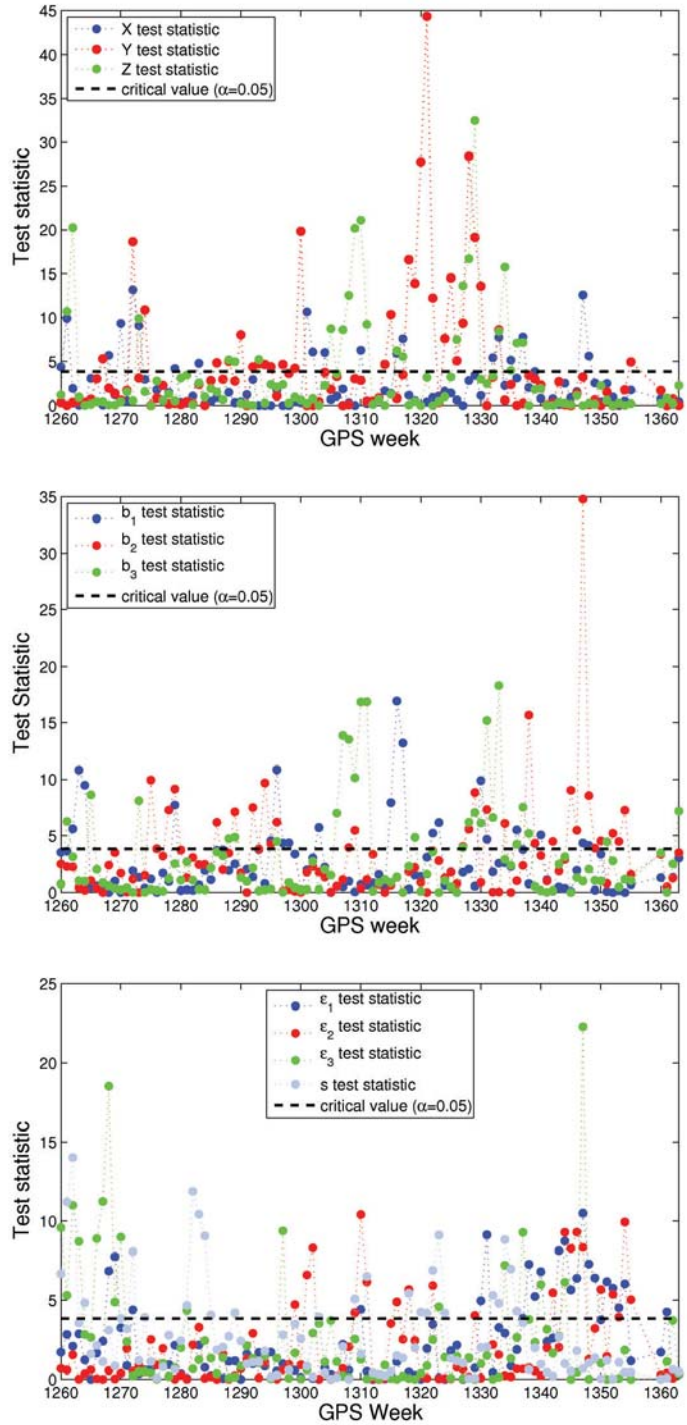


Fig. 8.19: Test for significance for the local parameters in the GPS/GRACE combination: geocentre motion X,Y,Z (top), Helmert transformation parameters $\underline{b}$ (middle), and the Helmert rotation parameters $\underline{\epsilon}$ and the Helmert scaling parameter s (bottom).

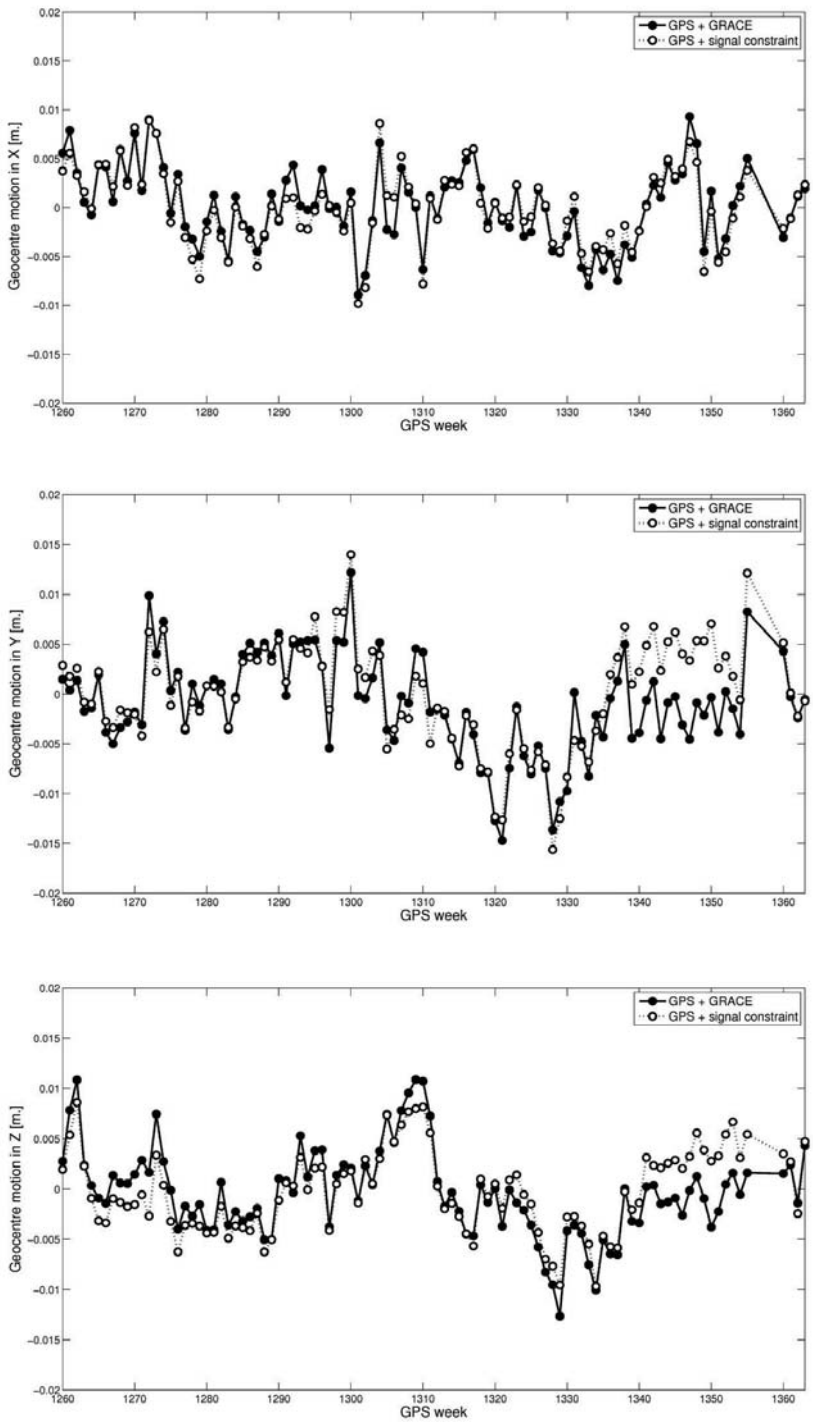


Fig. 8.20: Geocentre motion for each GPS week. The GPS-only solution (regularized by the signal constraint) is shown, together with the GPS/GRACE combined solution.

generate monthly gravity solutions. Using the Loading Love Numbers, these monthly models can be inverted into surface mass changes, expressed in equivalent water heights.

A different technique to estimate these surface mass changes is to invert deformations of the Earth, measured by GPS site displacements. The technique is quite accurate, but it lacks global coverage. Therefore, a GPS-only solution is unstable even for the very low degree spherical harmonics ($l > 3$). A solution to this problem is to regularize the solution. As the time-varying gravity signal should be very small over the oceans, we could damp the solution of the surface mass changes over the ocean. This can be viewed as a Tikhonov regularization. However, continental areas with only a few GPS sites, such as Northern Africa, still produce unstable results in these areas. We therefore modified the regularization procedure and damped those regions in which the temporal variations of the gravity signal were small (based on the GRACE estimates). The new regularization matrix improved the solution considerably, compared to the GRACE solutions.

Weekly GPS-only solutions ($L = 14$) could be obtained with the use of this regularization. The regularization parameter was determined using VCE techniques and was comparable to earlier estimates by independent ECCO solutions [Kusche and Schrama, 2005]. With a proper stochastic model, we could detect and treat possible outliers in the GPS site displacements. We have compared the conventional 3-sigma outlier removal technique to the robust M-estimation method (Huber, $k = 1.7$). The robust method only changes the solution close to the down-weighted observations, whereas the removal of the outliers destabilized the solution, changing the surface mass estimates around the globe. The importance of the covariances between the GPS site displacements was tested for the weekly GPS-only solutions. Large differences, up to half of the signal amplitude, could be seen when we compared this solution to the solution using full variance-covariance.

Due to inconsistencies between the GRACE models and GPS, we needed to add the regularization matrix in the estimation of the weights of the monthly GRACE model and the weekly GPS data sets, using VCE. These relative weights were then used to obtain the unregularized GPS/GRACE monthly solutions of the surface mass changes. We could see that the GPS data mainly contributed to the very low spherical harmonic coefficients, especially to degree 2.

As a by-product, weekly estimates of the motion of the geocentre could be obtained, together with 7 Helmert parameters to account for an inconsistency in the reference frame between the weekly GPS data sets and the monthly GRACE models. The geocentre parameters and the Helmert parameters were treated as local parameters and were tested for significance.

Future work should include satellite laser ranging data to improve the low-degree spherical harmonics and ocean bottom pressure data to stabilize the solution over the oceans. Moreover, we should extend the truncation degree and link the monthly solutions together, by using a singular value decomposition; see e.g., Schrama et al. [2007]. In this way, we can reduce the stripe pattern at the resonance terms of the CSR GRACE models and better separate signal from noise.

9

Application 3: Temporal aliasing of hydrological signals in a simulated GRACE recovery

One of the problems with the estimation of monthly gravity field solutions using GRACE observables is that short period mass variations alias into the monthly solutions. Errors in the ocean tide models and atmospheric models, as well as the variations of the hydrological (continental mass) signal within a month, will disturb the monthly mean gravity field estimates [Han et al., 2004].

In this study, we will try to reduce this temporal aliasing effect using VCE techniques. As the short period mass variations are not accounted for in the functional model (the reason for the aliasing), they will enter into the vector of residuals and should therefore be accounted for in the stochastic model. A reduction of the temporal aliasing effect is therefore possible by a refinement of the stochastic model by VCE.

The method is tested using a simulated GRACE orbit. The annual signal of the hydrological LaD model [Milly and Shmakin, 2002] is used as the simulated signal and the noise variance of GRACE is added to these simulated observations.

9.1 Temporal aliasing of signals

Before the GRACE observables can be inverted into the temporal (monthly) gravity solutions, one needs to subtract the influence of the ocean tide and the mass variations within the atmosphere. Comparisons between different ocean tide models, e.g., GOT99.2 [Ray, 1999], NAO99 [Matsumoto et al., 2000], and TPX06.2 [Egbert and Erofeeva, 2002] give an indication of the errors still present in the current ocean tide models. Although the estimation of monthly GRACE models will average out most of these errors, a part of the errors, especially the solar tides, will alias into the global

monthly GRACE solutions. This has been extensively discussed in Ray et al. [2003] and Seo et al. [2008].

The mass redistributions within the atmosphere generate a time-varying gravity signal, which can be modelled by surface pressure data, under the assumption that the atmosphere is condensed onto a very thin layer on the Earth's surface [Chao and Au, 1991]. The difference between the models of the European Center for Medium-range Weather Forecast (ECMWF) and the National Centers for Environmental Prediction (NCEP) is an indication of the error budget within these models, although the models are not completely independent of each other [Han et al., 2004]. In Velicogna et al. [2001] errors in the 6-hour forecast fields (NCEP) are used to predict the aliasing error in the monthly GRACE field due to the atmospheric correction.

One of the goals of the GRACE mission is to retrieve the temporal gravity field signal, which is caused by continental surface mass redistributions. However, this temporal signal will produce a temporal aliasing effect on its own, as short-term temporal variations will alias into the monthly solutions. In Jansen et al. [2008b], it was shown that sub-monthly GRACE solutions can be retrieved, which will partly reduce the temporal aliasing effect.

The surface water mass redistributions are caused by the hydrological cycle (precipitation, evaporation, transpiration, and run-off) and can be modelled by daily hydrological models. The GRACE mission will provide Monthly mean Water Storage Anomalies (MWSA), which is a relative signal, compared to a long-term mean. One can not distinguish between this long-term mean and the Earth's static field, i.e. the long-term mean of the hydrological signal will enter into the models of the static gravity field. In our test setup, we shall use Daily mean Water Storage Anomaly (DWSA) to identify the aliasing effect of this hydrological signal.

More information on the aliasing effect of errors in the ocean tide models, atmosphere and continental hydrology, together with a simulation study of the future GOCE mission, can be found in Han et al. [2006].

9.2 Test setup

If one wants to identify the effect of the temporal aliasing on the monthly GRACE solutions, one needs to work with a detailed orbital simulation of the GRACE mission, as the aliasing effect is also dependent on the orbit characteristics [Velicogna et al., 2001]. We simulated the gravity potential differences due to continental hydrology (expressed in spherical harmonics) between the two satellites GRACE-A $(\vartheta_A, \lambda_A, r_A)$ and GRACE-B $(\vartheta_B, \lambda_B, r_B)$ on the true GRACE orbits. The potential differences can be computed by

$$V^{AB}(\vartheta_A, \lambda_A, r_A, \vartheta_B, \lambda_B, r_B, t) = V(\vartheta_A, \lambda_A, r_A, t) - V(\vartheta_B, \lambda_B, r_B, t) \qquad (9.1)$$

with the scalar function $V(\vartheta, \lambda, r, t)$ computed by

$$V(\vartheta, \lambda, r, t) = \frac{GM}{R} \sum_{l=0}^{\infty} \left(\frac{R}{r}\right)^{l+1} \sum_{m=0}^{l} \left[\bar{C}_{lm}(t) \cos m\lambda + \bar{S}_{lm}(t) \sin m\lambda\right] \bar{P}_{lm}(\cos \vartheta). \tag{9.2}$$

Note that the input parameters, $\bar{C}_{lm}(t)$ and $\bar{S}_{lm}(t)$, are time-dependent. If we add noise, the functional model will read

$$\underline{y} = \underline{V}^{AB} + \underline{n} = A\underline{x} + \underline{e}, \tag{9.3}$$

with $\underline{n}$ the noise vector. The stochastic model of the residual vector $\underline{e}$ reads

$$E\{\underline{e}\} = \underline{0} \quad ; \quad D\{\underline{e}\} = Q_y. \tag{9.4}$$

As the vector of unknowns $\underline{x}$ contains the monthly mean spherical harmonic coefficients, $\bar{C}_{lm}$ and $\bar{S}_{lm}$, the temporal variations of these coefficients will be absorbed by the residual vector $\underline{e}$. We will try to use a stochastic model validation by VCE to reduce this aliasing effect.

9.2.1 Hydrological model

We have used the Land Dynamics (LaD) model [Milly and Shmakin, 2002], which is a model of land water and energy balance. The spatial resolution is $1^\circ \times 1^\circ$ grid, the temporal resolution is one month. From the monthly solutions, the annual signal (expressed in spherical harmonic coefficients) is retrieved. The solution was truncated to degree and order 60. We have limited ourselves to one month (August 2003) as a test case for our algorithms. For each day, a set of coefficients was generated from the annual signal. The mean signal for this particular month (August 2003) is subtracted from the daily solutions to obtain daily variations of the annual signal within one month. The variation of the annual signal, expressed in equivalent water heights, on the 1st of August 2003 with respect to the mean of August 2003 is drawn in figure 9.1. From these daily coefficients, potential differences (V^{AB}) between the two GRACE satellites are simulated along the orbit, with a sampling of 10 seconds.

9.2.2 Simulation of the noise

If we want to add the noise vector $\underline{n}$ to the temporal signal vector $\underline{V}^{12}$, we need a good estimation of the noise characteristics of the GRACE potential differences, which are the pseudo-observations. As the in situ observations from GRACE mainly come from the range-rate variations between the two satellites [Jekeli, 1999], the pseudo-observations V^{12} could be approximated by

$$V^{AB} \approx \|\dot{\underline{r}}_A\| \cdot \dot{\rho}_{AB}, \tag{9.5}$$

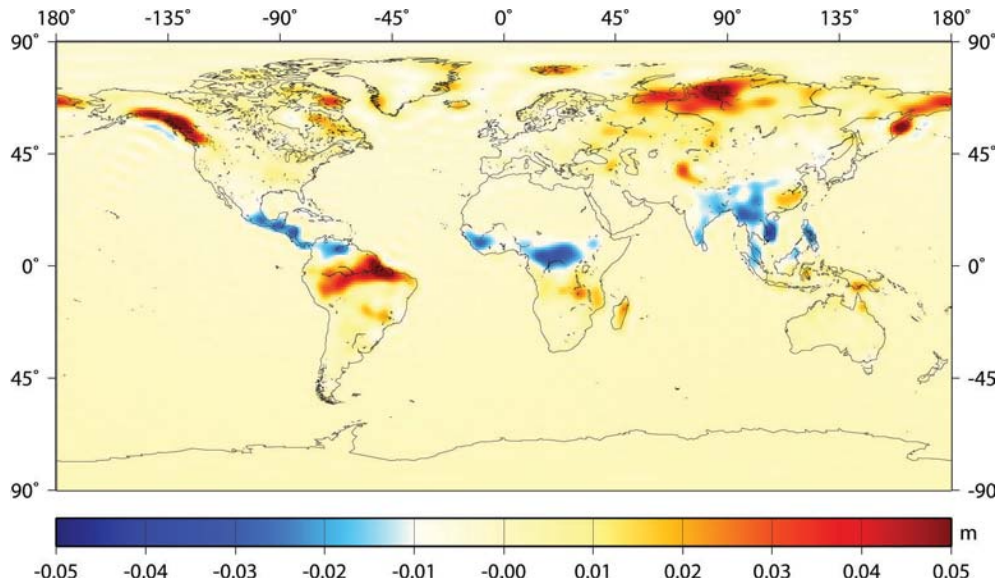


Fig. 9.1: Variation of the annual hydrological signal (LaD model) of day 1 (August 2003) with respect to the monthly mean, truncated at $L = 60$.

with ρ_{12} the range between between the two satellites measured by the highly accurate K-band ranging. From the propagation law of variances, we then obtain

$$\sigma_{V^{AB}} \approx \|\dot{\underline{r}}_A\| \cdot \sigma_{\dot{\rho}_{AB}} \tag{9.6}$$

As the post-processed range-rate data $\dot{\rho}_{AB}$ have an accuracy of approximately $0.1\ \mu\mathrm{m/s}$ [Han et al., 2004, 2005a] and the satellites fly at an along-track velocity of approximately $7.5\ \mathrm{km/s}$, the assumed standard deviation for the noise of the pseudo-observations in the energy balance approach is $\sigma_{V^{AB}} = 7.5 \cdot 10^{-4}\ \mathrm{m^2/s^2}$. We have simulated some variations along this value for different days.

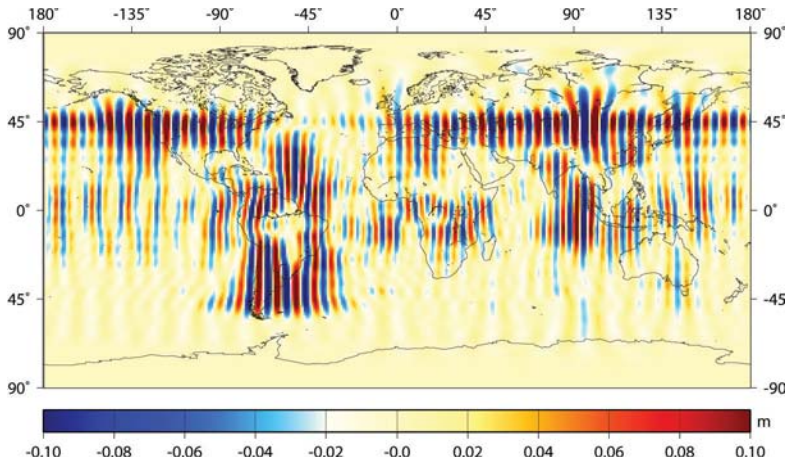
9.3 Results

We will perform three tests with the simulated data. First, we will only consider the hydrological signal, which gives us the opportunity to solely look at the temporal aliasing effect that comes from the daily variations of the annual signal. In the second test, we simulate white observation noise and try to estimate the characteristics of this noise by VCE. In the third test, we will combine the two components and look at the total contribution of the temporal aliasing effect and the noise.

9.3.1 The temporal aliasing effect

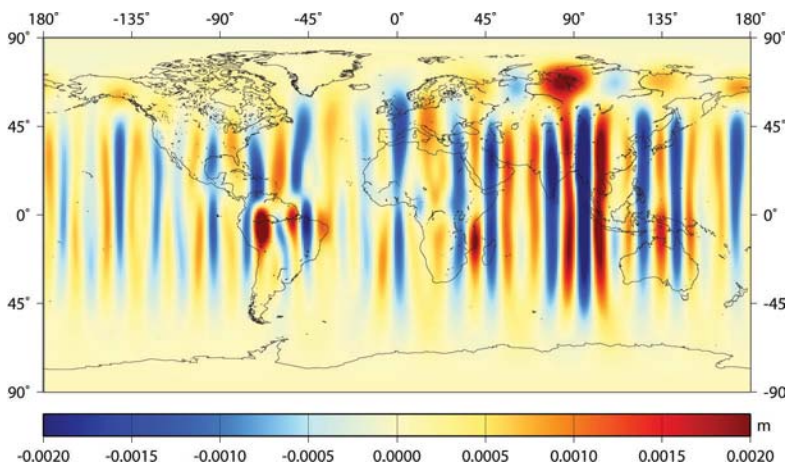
In this test, we will only consider the daily variations of the signal. We have generated 30 daily data sets, using the annual signal in the LaD model. From these 30 data sets,

normal equations are computed and solved for up to degree and order 60. The retrieved signal is the direct result of the temporal aliasing of the daily variations within one month. This signal is shown in figure 9.2.



Error statistics [m]

min.	-0.2454
max.	0.2233
mean	0.0000
RMS	0.0368



Error statistics [m]

min.	-0.0039
max.	0.0037
mean	0.0000
RMS	0.0007

Fig. 9.2: Temporal aliasing, expressed in terms of equivalent water heights [m], due to daily variations of the annual hydrological signal (August 2003), without a refinement of the stochastic model using VCE. Solution without Gaussian filtering (top), and a solution with Gaussian filtering (radius = 600 km) (bottom). Note the different color bars.

Although we used a Gaussian filtering with a radius of 600 kilometres (see section 8.2.2), we can still see a clear North-South pattern in the solution. Especially near areas where the daily variations show large fluctuations (see figure (9.1)), the solution is disturbed by the temporal aliasing effect.

As the temporal variations of the signal are not augmented in the functional model (in the estimation of monthly models), these variations will be absorbed by the residual vector $\underline{e}$ and therefore can not be separated from the observation noise. In this test setup, we set the noise to zero, making it possible to identify and quantify the temporal aliasing

effect. We have used the Iterative Restricted Maximum Likelihood Estimator (IREML) to estimate the stochastic model of the residual vector (and consequently the temporal aliasing effect). The estimated standard deviations per day of the month (Aug. 2003) are shown in figure 9.3.

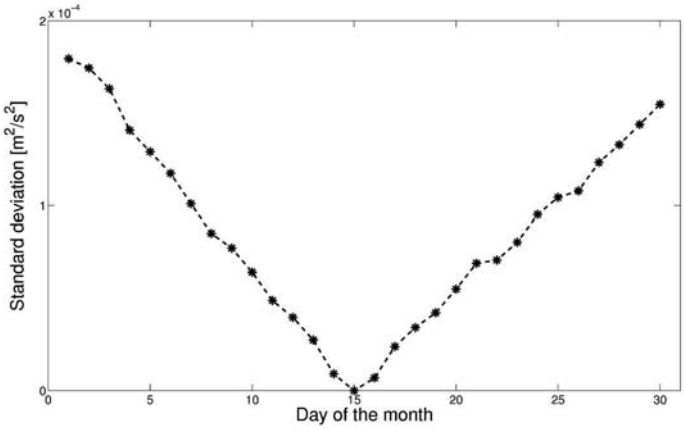


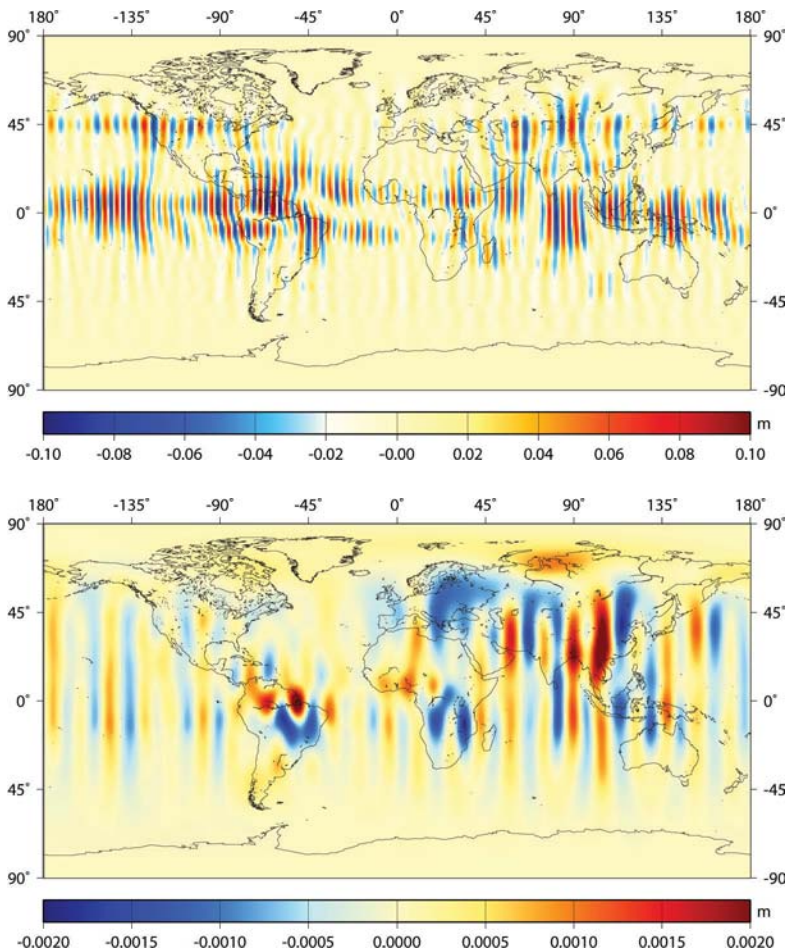
Fig. 9.3: Estimated standard deviations (by IREML) for August 2003 of the daily variations of the annual hydrological signal within one month.

As for each coefficient the monthly mean is subtracted from the observations, the days at the beginning and the end of a month show larger variations to this mean than the days along the middle of the month. This is also clearly visible in the estimated standard deviations. The data sets at the beginning and end of the month retrieve higher standard deviations than the other data sets. In Han [2004], it was shown that the daily variation of day 15 with respect to the monthly mean is below the GRACE noise level, already from degree 3, contrary to the daily variation of day 30, which is higher than the noise of the GRACE up to degree 17.

If we use this estimated stochastic model in a weighted least-squares approach, the data sets at the border will get less weight than the other data sets. This will reduce the temporal aliasing effect considerably. For this test case, the filtered results showed a decrease in RMS from 0.7 mm. to 0.5 mm, in terms of equivalent water heights. The improved solution is shown in figure 9.4.

9.3.2 The addition of random noise

Based on a-priori estimates of the error behaviour of the K-band range data [Han et al., 2004], we have come to an approximated noise level of $7.5 \cdot 10^{-4}$ $\mathrm{m}^2/\mathrm{s}^2$ for the potential differences between the two satellites; see also section 9.2.2. As the noise of the GRACE satellites is not heterogeneous, we have used different noise levels for each daily arc. First we will show the solution in which all data sets are weighted equally by taking a scaled identity matrix as the stochastic model. This solution is shown in figure 9.5.



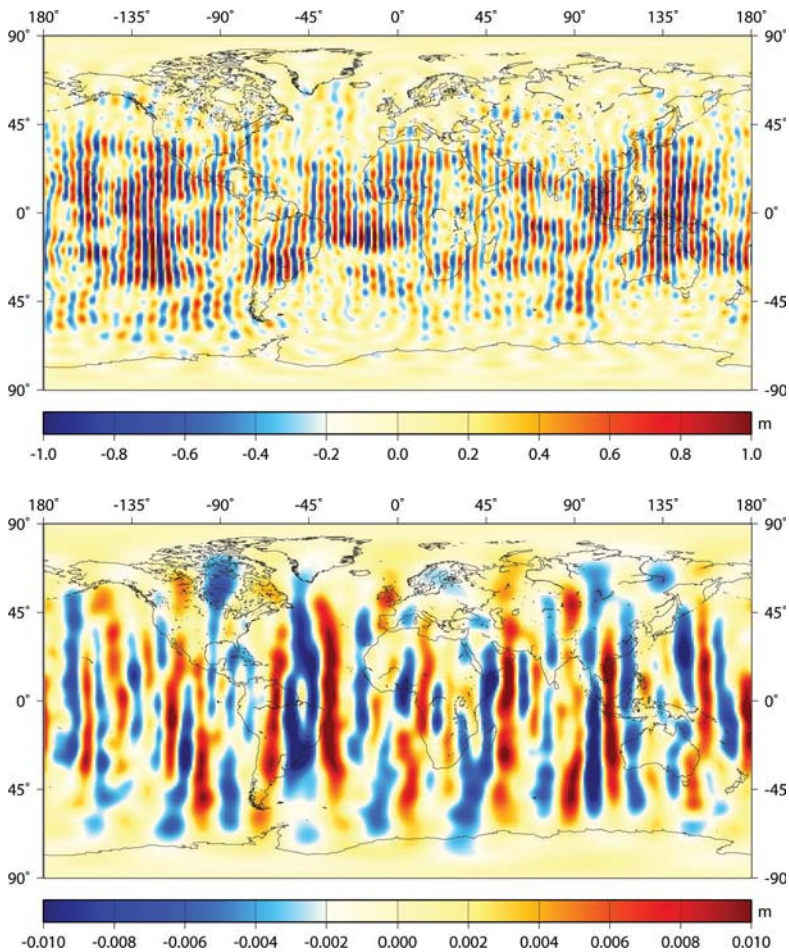
Error statistics [m]

min.	-0.1350
max.	0.1425
mean	0.0000
RMS	0.0221

Error statistics [m]

min.	-0.0019
max.	0.0027
mean	0.0000
RMS	0.0005

Fig. 9.4: Temporal aliasing errors, expressed in terms of equivalent water heights, due to daily variations of the annual hydrological signal (August 2003), after a refinement of the stochastic model using VCE. Solution without Gaussian filtering (top), and a Gaussian filtering (radius = $600\ \mathrm{km}$) (bottom). Note the different color bars.



Error statistics [m]

min.	-1.3886
max.	1.4376
mean	0.0000
RMS	0.3396

Error statistics [m]

min.	-0.0173
max.	0.0137
mean	0.0000
RMS	0.0042

Fig. 9.5: Propagation of the white noise in the simulated GRACE observations (potential differences) into a spherical harmonics model, complete to degree 60, without making use of VCE for stochastic model validation. Solution without Gaussian filtering (top), and with Gaussian filtering (radius = 600 km) (bottom). Note the different color bars.

The unfiltered solution shows equivalent water heights, which even exceed the 1 metre level, solely due to the white noise. The stochastic model and consequently the weighted least-squares solution was improved by VCE (IREML). The estimated standard deviations are found in figure 9.6, together with the estimated standard deviations of the temporal aliasing signal.

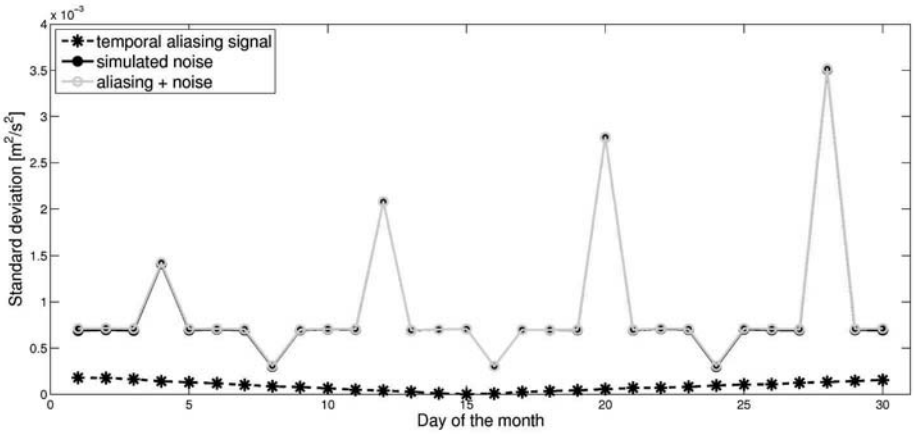


Fig. 9.6: Estimated standard deviations of the simulated white noise (by IREML), the temporal aliasing and their combined effect, for August 2003.

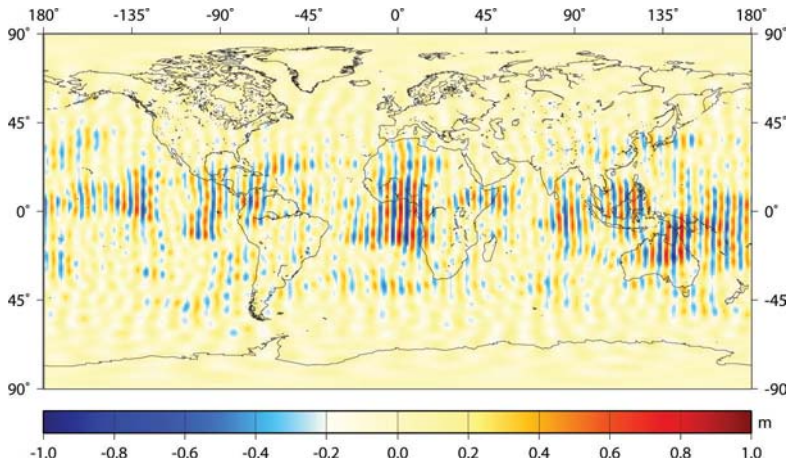
The estimated standard deviations are very close to the noise levels that were used as input in this test setup, a proof that the algorithms used produce unbiased estimates. When applying these standard deviations to the weighted least-squares solution, the effect of the noise can largely be reduced (figure 9.7).

9.3.3 The combination of noise and hydrological signal

The effect of the daily variations on the gravity field recovery is less than the measurement noise for the low-degree spherical harmonics ($l < 30$) [Han et al., 2004]. This is especially true for this test setup, as we only used a part of the hydrological signal. Therefore, the combination of both effects will produce nearly the same results as the noise-only solution; see figure 9.6 and table 9.1.

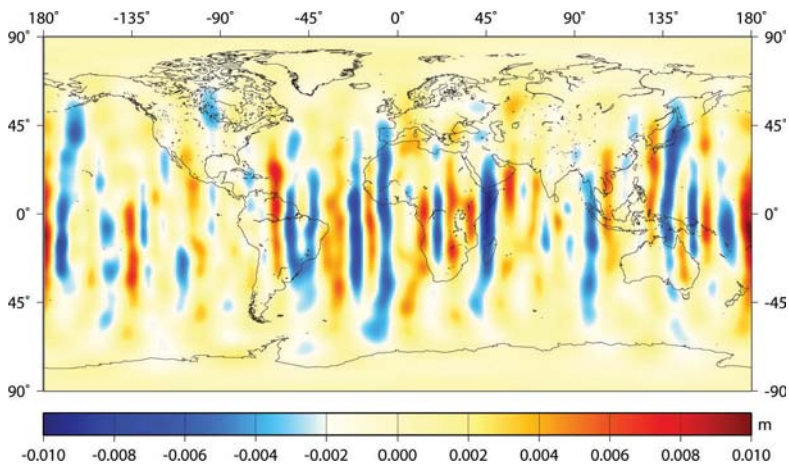
A way out of this is to incorporate more hydrological signal into the temporal aliasing data sets and to include error estimates of the ocean tide models and the atmospheric models. Han et al. [2004] mentioned that the daily variations within a month can lead to potential differences between the two satellites up to $1.0 \cdot 10^{-3}\ \mathrm{m}^2/\mathrm{s}^2$ above the continental areas, which is much larger than the estimated standard deviations estimated so far, see figure 9.3. With higher amplitudes of the daily variations of the hydrological signal, the temporal aliasing effect will become more visible in the combination of noise and the temporal signal.

We have written the stochastic model as a linear combination of daily cofactor matrices. This is not a realistic situation, as the properties of the observation noise will not



Error statistics [m]

min.	-0.9530
max.	0.9369
mean	0.0000
RMS	0.1928



Error statistics [m]

min.	-0.0104
max.	0.0107
mean	0.0000
RMS	0.0025

Fig. 9.7: Propagation of the white noise in the simulated GRACE observations (potential differences) into a spherical harmonics model, complete to degree 60, after a proper stochastic model validation using VCE. Solution without Gaussian filtering (top), and with Gaussian filtering (radius = 600 km) (bottom). Note the different color bars.

Tab. 9.1: Propagation of the temporal aliasing effect, simulated white noise and a combination of both, on a spherical harmonic solution ($L = 60$), expressed in equivalent water heights [m]. No Gaussian filtering has been applied.

		Temp. aliasing	Noise-only	Combination
no VCE	Max.	0.245	1.438	1.440
	RMS	0.037	0.340	0.343
VCE	Max.	0.143	0.953	0.991
	RMS	0.022	0.193	0.207

change on a daily basis. Moreover, one should incorporate the noise covariances in the stochastic model and include spatial information of the aliasing effect in the stochastic model. This will be the scope of future studies on temporal aliasing.

9.4 Summary and outlook

With the launch of the GRACE satellites, it is not only possible to estimate the static gravity field of the Earth, but also temporal deviations from this static field are estimated on a monthly basis. The contributions of the ocean tides and mass changes in the atmosphere should be modelled and subtracted from the observations in order to separate them from the gravity field solutions. Errors in these models will partly alias into the monthly GRACE solutions. Moreover, the hydrological variations within a month are not augmented in the functional model and will therefore alias into these monthly solutions.

As the ocean tide errors, atmospheric model errors and the hydrological variations (within a month) are not augmented in the functional model, they will be partly absorbed by the vector of residuals. The part that is not absorbed by the vector of residuals will alias into the solution. As the weighted least-squares solution will minimize the residual square sum, we need to adjust our stochastic model to account for these errors (next to the measurement noise). We will use variance component estimation to validate the stochastic model and in this way try to down-weight the data sets with large errors in the ocean tide model, the atmospheric model or with large hydrological variations, compared to the monthly mean.

In our test setup, we have used the annual signal of the hydrological LaD model and simulated pseudo-observations (potential differences) along a real GRACE orbit. We have divided the data set into daily arcs and estimated variance components for each daily arc. In a noise-free setting, the arcs in the middle of the month obtained much more weight (VCE) than the arcs at the beginning and end of the month. The hydrological variations with respect to the monthly mean (temporal aliasing effect) are much higher at the boundaries of a month than close to the middle of the month. So far, the VCE looks promising in reducing the temporal aliasing effect in monthly GRACE solutions.

However, if we add noise to the observations, we can see that in this test setup, the simulated noise of GRACE is much higher than the generated temporal aliasing signal. Variance component estimation mainly followed the noise levels of the different daily arcs.

In the future, we need to simulate more temporal aliasing signal by including ocean tide errors, atmospheric model errors and more hydrological signal than the annual signal. Moreover, the GRACE data sets should be divided into smaller arcs (than the daily arcs) and noise covariances should be incorporated in the stochastic model of GRACE.

10

Application 4: The computation of a height reference surface in Switzerland

As was stated earlier in section 2.1.5, one can not simply use the gravimetric geoid to derive orthometric heights H from ellipsoidal heights h, which are mostly measured by GPS. The discrepancy $\delta h_i = h_i - N_i - H_i$ is caused by many errors, including long-wavelength errors in the gravimetric geoid, and systematic and stochastic errors in the levelling data and the GPS data.

A solution to this problem is to estimate a corrector surface, which best fits these discrepancies, and to add this to the gravimetric geoid. In this way, one obtains a height reference surface, which can be used as a transformation surface between the measured GPS measurements and the desired orthometric heights.

In this study, we will compute such a height reference surface for Switzerland. The corrector surface is computed using a combination of gravimetric, GPS and levelling data sets, each with full variance-covariance information. Variance component estimation is used to re-weight the different data sets and a robust M-estimation technique is used to down-weight possible outliers in the data sets. Different functional models to estimate the corrector surface are compared to each other and the parameters are tested for significance. It will be shown that the parametrization of the corrector surface has an effect on the VCE, the quality description of the variance components, and the treatment of the outliers.

10.1 Test setup

The old height reference surface of Switzerland, LN02, was purely based on levelled height differences, without making use of information on the gravitational field of the

Earth. Moreover, it did not take the uplift of the Alps into account, which is of the order of $1.5\ \mathrm{mm/year}$ with respect to the Central Plateau [Schlatter and Marti, 2002]. A new height reference surface was needed, which was called LHN95 [Schlatter, 2006]. The orthometric heights of LHN95, which take the uplift of the Alps and the gravitational field into account, show differences with LN02 up to several decimeters. The most recent national geoid model is called CHGeo2004, which is a combination of the gravimetric solution, astro-geodetic data and GPS/levelling data [Marti, 2007]. As the GPS/levelling data receive high weights, CHGeo2004 should be considered as a height reference surface for Switzerland. Levelling measurements (LHN95), GPS measurements (LV95 campaign) and the CHGeo2004 surface are consistent with each other [Marti and Schlatter, 2005].

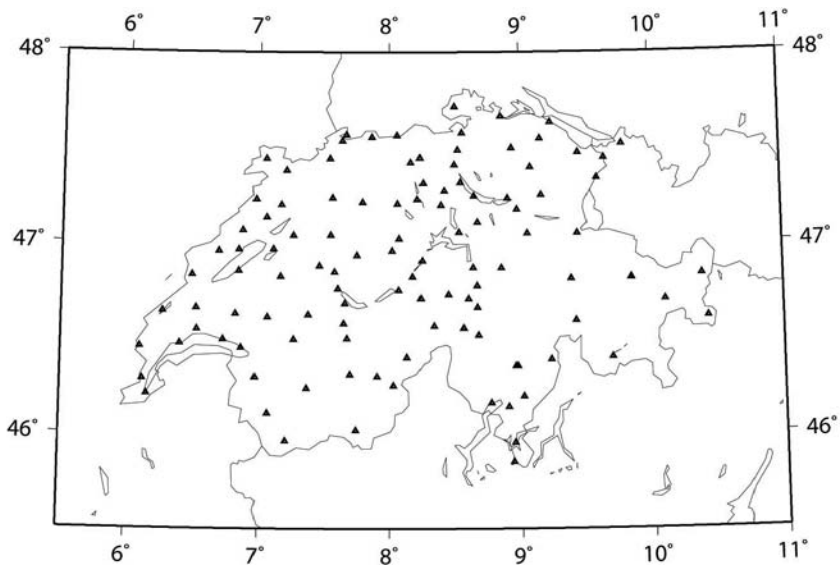


Fig. 10.1: Available 111 GPS/levelling points in Switzerland.

We will use a data set of 111 GPS/levelling points (figure 10.1) with the corresponding gravimetric geoid estimates to model the correction surface for the gravimetric geoid of Switzerland. The data set, covering the entire country of Switzerland, was kindly provided by Urs Marti from the Federal Office of Topography Swisstopo, together with full variance-covariance information on the GPS observations, the levelling data, and the gravimetric solution.

If we write the parametrization of the corrector surface as $A\underline{x}$, the functional model reads

$$\underline{y} = \underline{h} - \underline{N} - \underline{H} = A\underline{x} + \underline{e}, \tag{10.1}$$

with the corresponding stochastic model defined by

$$E\{\underline{e}\} = \underline{0} \quad ; \quad D\{\underline{e}\} = Q_y = \gamma_h Q_h + \gamma_N Q_N + \gamma_H Q_H, \tag{10.2}$$

where Q_h, Q_N, and Q_H are the cofactor matrices of respectively the GPS measurements $\underline{h}$, the gravimetric data $\underline{N}$ and the levelling data $\underline{H}$.

10.1.1 The data sets used

The GPS measurements and their variance-covariance matrix are the output of a commercial post-processing software package [Fotopoulos, 2005]. The correlations between neighbouring points are much lower than in the stochastic models of the orthometric heights and the gravimetric geoid [Marti et al., 2002].

The orthometric heights of LHN95 are computed from the adjustment of the national first- and second-order levelling measurements [Schlatter, 2006]. As the VC-matrix is the direct result from the adjustment of the levelling measurements, the orthometric heights between neighbouring points are highly correlated [Marti et al., 2002].

The gravimetric geoid was obtained by least-squares collocation. The variance-covariance information in the GPS/levelling points is the outcome of an error propagation procedure. As the uncertainty contribution of the global geopotential model is not incorporated into the VC-matrix, the a-priori standard deviations of the gravimetric geoid estimates are too optimistic. Nevertheless, the largest (a-priori) uncertainty in the pseudo-observations δh_i comes from the gravimetric solution, with an average standard deviation of $1.93\ \mathrm{cm}$.

10.1.2 The four-parameter corrector surface

Many suggestions have been made in literature to define a parametric model for the corrector surface. De Min [1996] uses a simple tilted plane to correct the gravimetric geoid of the Netherlands to the height reference surface. One could extent this to a parametric model with an increasing number of free parameters, up to the number of observations (= number of GPS-levelling points). One should however be cautious and test the parameters for significance.

A well-known parametrization is the four-parameter form, which accounts for a datum inconsistency between the GPS datum and the datum used to estimate the gravimetric geoid. In Fotopoulos [2005], this form is used for exactly the same Swiss data set, as is used in this study. The functional model of observation i then reads

$$f(\phi_i, \lambda_i) = x_1 + \cos\phi_i \cos\lambda_i \cdot x_2 + \cos\phi_i \sin\lambda_i \cdot x_3 + \sin\phi_i \cdot x_4, \tag{10.3}$$

with x_j the four unknown parameters. We will use this parametric model to test our algorithms. In section 10.4, several parametric models will be compared to each other.

As a first estimate, we set $\gamma_h = \gamma_N = \gamma_H = 1.0$ and leave all outliers as they are. The estimated corrector surface is shown in figure 10.2. Based on the vector of residuals $\underline{e}$ and the stochastic model Q_y, we have computed the standardized residuals. Although one observation had a test statistic above 3.0, we did not perform any outlier treatment at this stage.

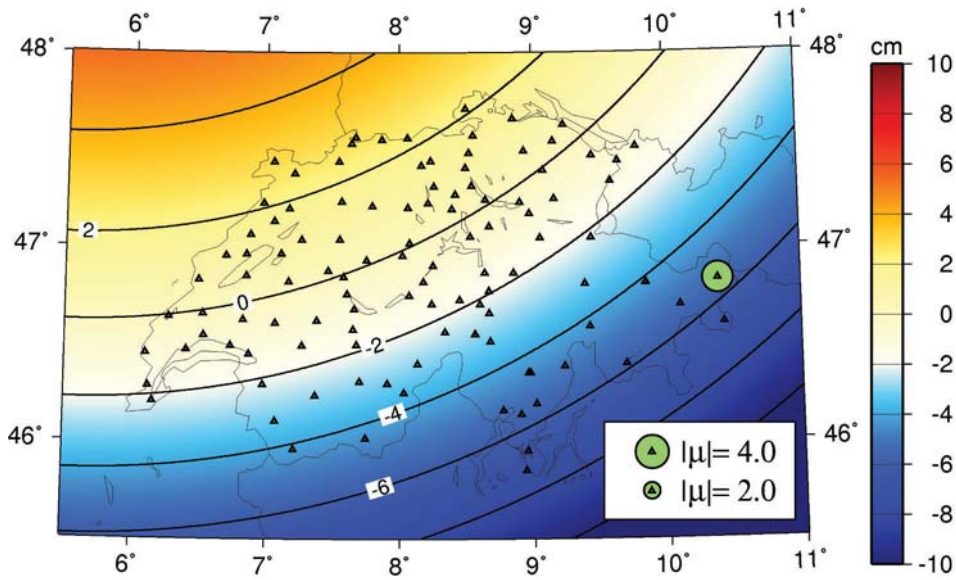


Fig. 10.2: Correction surface, using the a-priori stochastic model. The green circle indicate the test statistic of the observation with a test statistic above 3.0.

10.2 Validation of the stochastic model

The estimation of the vector of unknowns $\underline{x}$ assumes a proper stochastic model. This stochastic model will be estimated by VCE by a rescaling of the cofactor matrices. Moreover, the stochastic model will propagate into the quality estimate of $\underline{x}$, it will be used in the treatment of outliers, and in the tests for significance of the parameters of the corrector surface.

10.2.1 Variance Component Estimation

We will use the general form of the MINQUE, i.e. Eq. (4.28), to estimate the variance components of the cofactor matrices and its error estimates. If we assume the observations to be normally distributed, we can obtain a quality description of the estimated variance components. If we do not treat the outliers and we assume full variance-covariance information, the estimated variance components and their VC-matrices have the values

$$\underline{\hat{\gamma}} = \begin{bmatrix} \hat{\gamma}_h \\ \hat{\gamma}_N \\ \hat{\gamma}_H \end{bmatrix} = \begin{bmatrix} 2.74 \\ 0.98 \\ 4.04 \end{bmatrix} \quad ; \quad Q_{\hat{\gamma}} = \begin{bmatrix} 1.49 & -0.19 & -0.72 \\ -0.19 & 0.11 & -0.38 \\ -0.72 & -0.38 & 8.03 \end{bmatrix}. \tag{10.4}$$

The results were quite similar to Fotopoulos [2005], who performed the same study. The estimated corrector surface, together with its difference from the previous estimate (using equal weighting) is shown in figure 10.3.

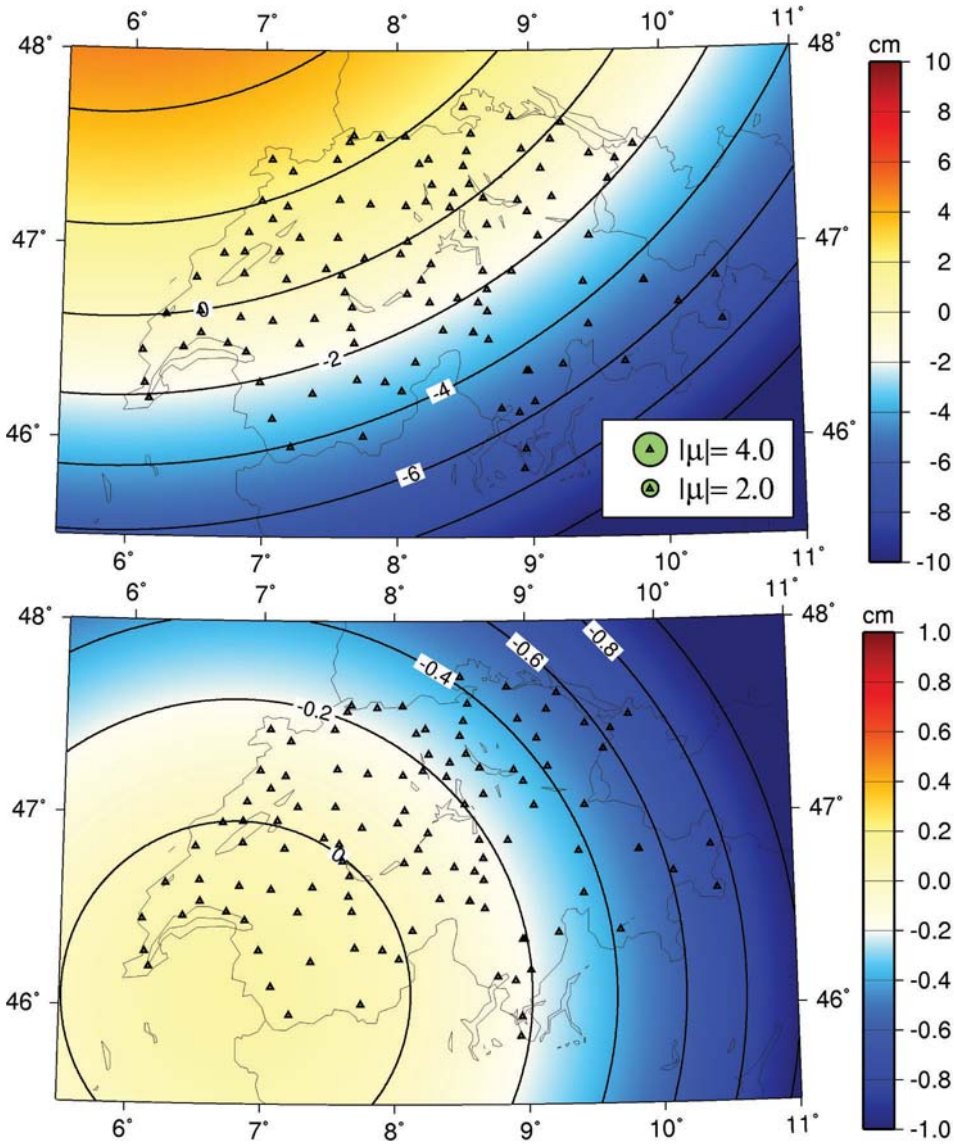


Fig. 10.3: Top: corrector surface, using MINQUE to estimate the stochastic model. Bottom: Differences with respect to the model using the a-priori stochastic model.

10.2.2 Diagonalized variance-covariance matrices

If we would neglect the covariances of the VC-matrices, this would evidently have an effect on the estimation of the variance components and consequently the corrector surface. The MINQUE estimates at convergence are

$$\underline{\hat{\gamma}} = \begin{bmatrix} \hat{\gamma}_h \\ \hat{\gamma}_N \\ \hat{\gamma}_H \end{bmatrix} = \begin{bmatrix} 0.82 \\ 1.03 \\ 3.65 \end{bmatrix} \quad ; \quad Q_{\hat{\gamma}} = \begin{bmatrix} 2.49 & -0.11 & -2.58 \\ -0.11 & 0.20 & -0.93 \\ -2.58 & -0.93 & 9.88 \end{bmatrix}, \tag{10.5}$$

and are also quite similar to Fotopoulos [2005]. The uncertainties of the variance components have increased significantly, as we would expect. Moreover, the correlations between the estimates of the different variance components increased. If we define the correlation as

$$\rho_{ij} = \frac{Q_{ij}}{\sqrt{Q_{ii} \cdot Q_{jj}}}, \tag{10.6}$$

the correlation between the ellipsoidal variance component $\hat{\underline{\gamma}}_h$ and the orthometric variance component $\hat{\underline{\gamma}}_H$ changed from -0.21 to -0.52, when assuming diagonal cofactor matrices. The effect of this alternative stochastic model on the estimation of the corrector surface is shown in figure 10.4. Hence, neglecting the covariances in the observations can lead to changes in the corrector surface up to 1.5 cm.

10.3 Treatment of outliers

With the new stochastic model, with full variance-covariance information and MINQUE to re-scale the cofactor matrices, we can detect and treat possible outliers. We will compare the conventional 3-sigma rule with the robust M-estimation technique (Huber).

10.3.1 Conventional 3-sigma rule

Under the consideration of the new stochastic model, we did not find any observation with a test statistic above 3.0. Therefore, no outliers were removed, according to the conventional 3-sigma rule.

10.3.2 Robust M-estimation

As the conventional 3-sigma rule did not find any possible outliers, we can state that there are no large outliers in the data. However, erroneous observations with a test statistic between 1.7 and 3.0 can have a large effect on the estimation of the corrector surface, if the observation is far from any other observation. The robust M-estimation down-weights 6 observations within the data set of 111 observations; see figure 10.5. Especially the observations at the borders of the Swiss data set can have a large influence on the estimation of the corrector surface. The down-weighting of the 6 observations

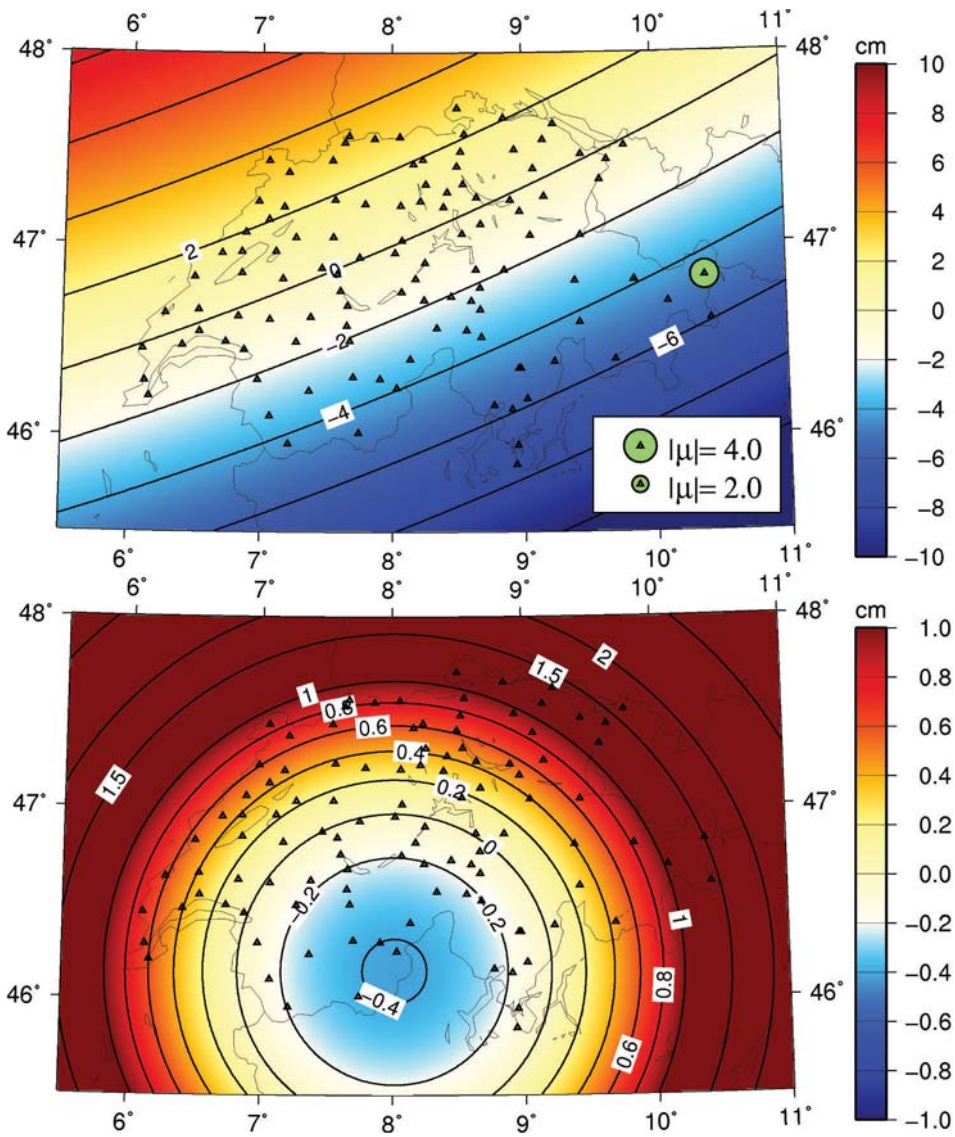


Fig. 10.4: Top figure: corrector surface, using MINQUE to estimate the stochastic model, with diagonal cofactor matrices. The bottom figure shows the differences with the MINQUE solution, using full cofactor matrices.

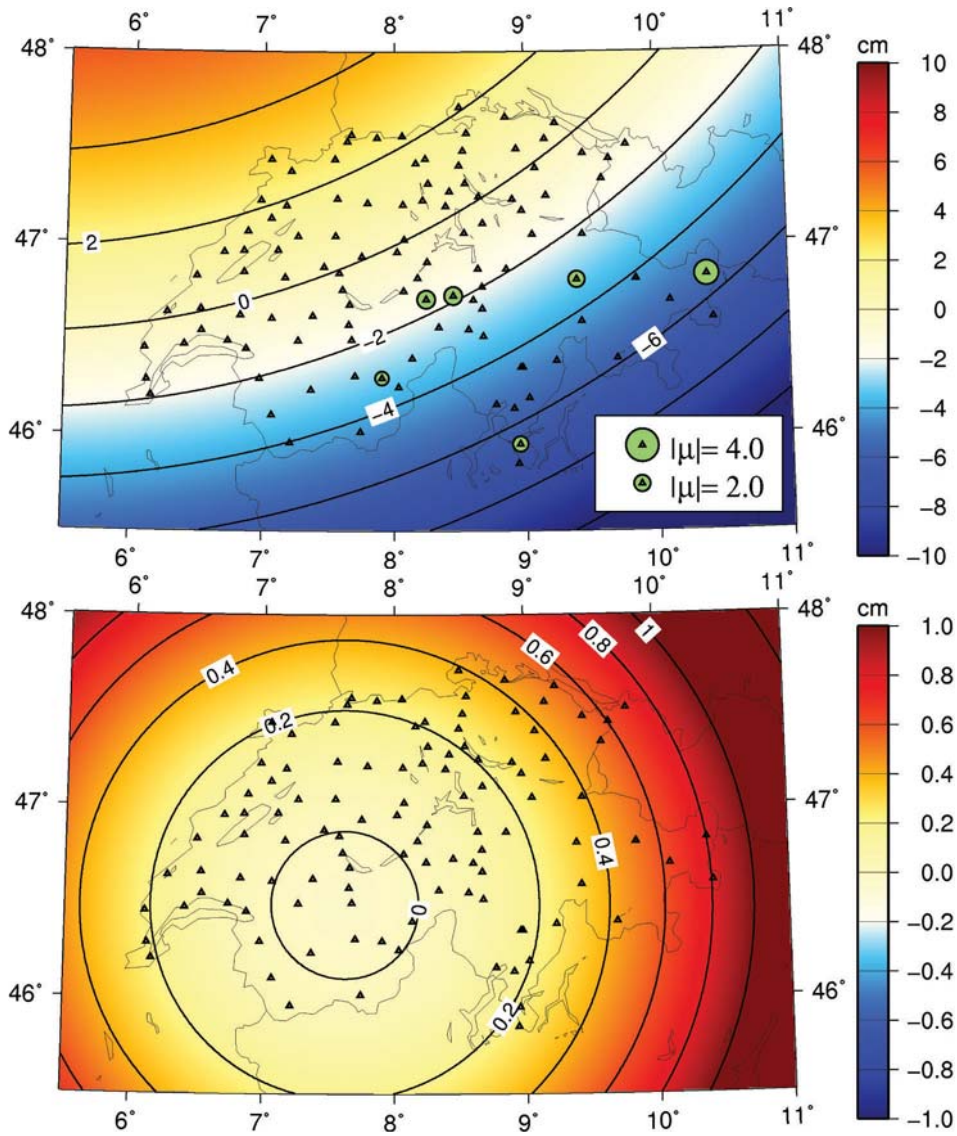


Fig. 10.5: Top figure: corrector surface, using MINQUE to estimate the stochastic model and the robust M-estimation (Huber, $k = 1.7$) to down-weight possible outliers. The green circles indicate the estimated test statistic. The bottom figure shows the differences with the MINQUE solution, without any outlier treatment.

changed the corrector surface up to 1 cm. Moreover, the M-estimation changed the variance components and their VC-matrix, i.e.,

$$\underline{\hat{\gamma}} = \begin{bmatrix} \hat{\gamma}_h \\ \hat{\gamma}_N \\ \hat{\gamma}_H \end{bmatrix} = \begin{bmatrix} 2.90 \\ 1.17 \\ 4.01 \end{bmatrix} \quad ; \quad Q_{\hat{\gamma}} = \begin{bmatrix} 1.75 & -0.22 & -0.82 \\ -0.22 & 0.14 & -0.46 \\ -0.82 & -0.46 & 9.14 \end{bmatrix}. \tag{10.7}$$

As the outliers are now properly treated (M-estimation) and the variance-covariance matrices are re-scaled by VCE, we can test for significance of the parameters of the corrector surface (Eq. (3.41)). Tests for significance of the parameters of the corrector surface show that all 4 parameters are insignificant, with test statistics $\delta\hat{\Omega}$ ranging from 0.11 to 0.50. These low test statistics imply are incorrect parametrization of the corrector surface. In the next section, we will compare several alternative parametrizations of this corrector surface.

10.4 Other parametrizations of the corrector surface

In this section, we will try to find a suitable parametrization of the corrector surface. As we need residuals for the outlier treatment and the VCE, we will not consider interpolation methods, like the multiquadric interpolation [Hardy, 1971]. One class of possible parametrizations is the class of polynomials of the form

$$p_N(\phi, \lambda) = \sum_{i=0}^{N} \sum_{j=0}^{N} p_{ij} \cdot \phi^i \cdot \lambda^j \cdot \cos^j \phi, \tag{10.8}$$

see e.g., Jiang and Duquenne [1996]. Additionally, we will account for the effect of a datum transformation on the geoid height N, i.e.,

$$q(\phi, \lambda) = \Delta a + \Delta X \cos\phi \cos\lambda + \Delta Y \cos\phi \sin\lambda + \Delta Z \sin\phi + a\Delta f \sin^2\phi, \tag{10.9}$$

with ΔX, ΔY, ΔZ the shift parameters between two datums, Δf the change in the flattening of the ellipsoid, and Δa the change in the semi-major axis a of the ellipsoid; see e.g., Heiskanen and Moritz [1967] and Kotsakis and Sideris [1999]. If we add $q(\phi, \lambda)$ to $p_N(\phi, \lambda)$, we get the parametrization

$$f_N(\phi, \lambda) = p_N(\phi, \lambda) + q(\phi, \lambda). \tag{10.10}$$

As p_{00} and Δa will both account for a bias in the discrepancies δh_i, we will only consider p_{00}.

One should be cautious to add extra parameters to the functional model as numerical instabilities can occur by an over-parametrization of the corrector surface [Fotopoulos, 2005]. The parameters have therefore been tested for significance (Eq. (3.41)). This is done in an iterative way. If multiple parameters fail the test of significance, we first

remove the parameter that has the lowest test statistic from the functional model. The iteration stops when all remaining parameters pass the test of significance. The resulting parametrization f'_N after the elimination of the insignificant parameters is then compared to the other parametrizations.

We need to define tests to compare the different parametrizations with each other. An indication of the goodness of fit are the standard deviations of the estimated variance components. Low standard deviations of the variance components imply that the observations are consistent with the stochastic and functional model.

Additionally, one could make a comparison by testing for the significance of the full vector of unknowns, i.e., $\hat{\Omega} = \underline{\hat{x}} N \underline{\hat{x}}$. As this test statistic largely depends on the number of parameters, it is better to look at the probability in the right-hand tail that the vector of unknowns is not significant. This probability depends on $\hat{\Omega}$ and the number of parameter n. In table 10.1, the results are shown for the different parametrizations.

Parametrization up to $N = 0$

We start with $N = 0$, which implies the estimation of a bias, if we only consider $p_0(\phi, \lambda)$:

$$p_0(\phi, \lambda) = p_{00}. \tag{10.11}$$

As was also found by Fotopoulos [2005], this parametrization results in a high estimate of the variance component for the levelling observations γ_H. This is probably due to systematic errors (e.g., tilt) in these observations, which will, if not properly treated in the functional model, be absorbed in the vector of residuals. The test for significance for the bias in this parametrization had the value of $\alpha = 0.07$, i.e. test statistic was lower than the threshold $\kappa_{0.05}$, which indicates that this parametrization is not significant.

The addition of $q(\phi, \lambda)$ increased the number of parameters to 5. With this parametrization, one estimates a datum transformation between two ellipsoids. However, two parameters failed the (iterative) test for significance, resulting in the parametrization

$$f'_0(\phi, \lambda) = \Delta X \cos\phi \cos\lambda + \Delta Z \sin\phi + a\Delta f \sin^2\phi, \tag{10.12}$$

Parametrization up to $N = 1$

The polynomial parametrization up to $N = 1$ leads to the function

$$p_1(\phi, \lambda) = p_{00} + p_{10} \cdot \phi + p_{01} \cdot \lambda \cdot \cos\phi + p_{11} \cdot \phi \cdot \lambda \cdot \cos\phi. \tag{10.13}$$

However, the parameter p_{11} of the basis function $\phi \cdot \lambda \cdot \cos\phi$ fails the test of significance. Removing this parameter from the parametrization results in the functional model of a tilted plane. The results of a tilted plane are quite good, compared to $p_1(\phi, \lambda)$ and the four parameter form; see table 10.1. The addition of $q(\phi, \lambda)$ and the elimination of the parameters that failed the test for significance, results in the function $f'_1(\phi, \lambda)$:

$$f'_1(\phi, \lambda) = p_{01} \cdot \lambda \cdot \cos\phi + p_{11} \cdot \phi \cdot \lambda \cdot \cos\phi + \Delta Y \cos\phi \sin\lambda. \tag{10.14}$$

Tab. 10.1: Variance components, their standard deviations and a test for significance of the vector of unknowns for different parametrizations of the corrector surface.

	n	$\hat{\gamma}_h$	$\hat{\sigma}_{\gamma_h}$	$\hat{\gamma}_N$	$\hat{\sigma}_{\gamma_N}$	$\hat{\gamma}_H$	$\hat{\sigma}_{\gamma_H}$	α
Four parameter form	4	2.90	1.32	1.17	0.38	4.01	3.02	$2.33 \cdot 10^{-5}$
$p_0(\phi, \lambda)$ (bias)	1	2.71	1.38	1.12	0.41	9.72	4.51	$6.91 \cdot 10^{-2}$
$f_0'(\phi, \lambda)$	3	3.00	1.34	1.17	0.38	3.73	2.88	$8.42 \cdot 10^{-6}$
$p_1(\phi, \lambda)$	4	2.93	1.34	1.20	0.38	4.04	3.03	$5.37 \cdot 10^{-5}$
$p_1'(\phi, \lambda)$ (tilted plane)	3	2.95	1.33	1.18	0.38	3.90	2.93	$1.32 \cdot 10^{-5}$
$f_1'(\phi, \lambda)$	3	3.01	1.33	1.14	0.37	3.61	2.81	$2.61 \cdot 10^{-6}$
$p_2(\phi, \lambda)$	9	2.85	1.29	1.07	0.36	3.42	2.85	$4.44 \cdot 10^{-6}$
$p_2'(\phi, \lambda)$	3	2.99	1.33	1.15	0.37	3.69	2.85	$4.27 \cdot 10^{-6}$
$f_2'(\phi, \lambda)$	4	3.29	1.37	0.99	0.35	3.73	2.81	$8.66 \cdot 10^{-7}$
$p_3'(\phi, \lambda)$	4	3.36	1.41	0.99	0.36	4.33	3.04	$7.36 \cdot 10^{-6}$
$f_3'(\phi, \lambda)$	3	3.01	1.33	1.16	0.37	3.61	2.83	$4.11 \cdot 10^{-6}$

Parametrization up to $N = 2$

The polynomial representation up to $N = 2$ can be written as

$$p_2(\phi, \lambda) = \sum_{i=0}^{2} \sum_{j=0}^{2} p_{ij} \cdot \phi^i \cdot \lambda^j \cdot \cos^j \phi. \tag{10.15}$$

As the number of parameters is increased to 9, this functional model has a better fit to the data than any of the previous parametrizations. Therefore, the estimated variance components show the lowest results; see table 10.1. However, many of the parameters fail the test of significance. The function $p_2(\phi, \lambda)$ is therefore not a good parametrization of the corrector surface.

The addition of $q(\phi, \lambda)$ to $p_2(\phi, \lambda)$ results in a numerical unstable system of normal equations, as some basis functions are almost identical to each other. We solve the problem by first removing the parameters p_{ij} that failed the test of significance:

$$p_2'(\phi, \lambda) = p_{01} \cdot \lambda \cdot \cos \phi + p_{11} \cdot \phi \cdot \lambda \cdot \cos \phi + p_{12} \cdot \phi \cdot \lambda^2 \cdot \cos^2 \phi. \tag{10.16}$$

We then add $q(\phi, \lambda)$ and test for significance of the parameters. We obtain

$$f_2'(\phi, \lambda) = p_{01} \cdot \lambda \cdot \cos \phi + p_{12} \cdot \phi \cdot \lambda^2 \cdot \cos^2 \phi + \Delta X \cdot \cos \phi \cdot \cos \lambda + \Delta Y \cdot \cos \phi \cdot \sin \lambda \tag{10.17}$$

Parametrization up to $N = 3$

The series of polynomials up to $N = 3$ resulted in an unstable system of normal equations. Some basis functions were too close to each other for this small area and the low amount of observations. After stabilizing the solution by removing some parameters we

tested the parameters for significance. The parametrization with significant parameters reads

$$p'_3(\phi,\lambda) = p_{20}\cdot\phi^2 + p_{21}\cdot\phi^2\cdot\lambda\cdot\cos\phi + p_{22}\cdot\phi^2\cdot\lambda^2\cdot\cos^2\phi + p_{13}\cdot\phi\cdot\lambda^3\cdot\cos^3\phi. \quad (10.18)$$

The addition of $q(\phi,\lambda)$ and subsequently the removal of those parameters that failed the test of significance, resulted in the function $f'_3(\phi,\lambda)$:

$$f'_3(\phi,\lambda) = p_{00} + p_{13}\cdot\phi\cdot\lambda^3\cdot\cos^3\phi + \Delta X\cdot\cos\phi\cdot\cos\lambda \quad (10.19)$$

We can further extend the parametrizations to $N > 4$, but this will over-parameterize the functional model [Fotopoulos, 2005].

The corrector surface

If we look at the test for significance (table 10.1), we can see that both the $f'_1(\phi,\lambda)$ and $f'_2(\phi,\lambda)$ parametrizations have very low probability α, i.e., the probability that the vector of unknowns is not significant is very low. A comparison between the functions shows that the estimates of the variance components of the $f'_1(\phi,\lambda)$ function are closer to the estimates of the other functions, compared to the $f'_2(\phi,\lambda)$. We therefore have chosen to use this parametrization of the corrector surface, i.e.,

$$f'_1(\phi,\lambda) = p_{01}\cdot\lambda\cdot\cos\phi + p_{11}\cdot\phi\cdot\lambda\cdot\cos\phi + \Delta Y\cos\phi\sin\lambda,$$

although some other parametrizations, mentioned in table 10.1, will be suitable as well. The chosen parametrization, together with a robust treatment of the outliers and MINQUE to scale the cofactor matrices, leads to the corrector surface shown in figure 10.6. The estimated variance components and their VC-matrix changed slightly, compared to the four parameter form:

$$\underline{\hat{\gamma}} = \begin{bmatrix} \hat{\gamma}_h \\ \hat{\gamma}_N \\ \hat{\gamma}_H \end{bmatrix} = \begin{bmatrix} 3.01 \\ 1.14 \\ 3.61 \end{bmatrix} \quad ; \quad Q_{\hat{\gamma}} = \begin{bmatrix} 1.77 & -0.23 & -0.76 \\ -0.23 & 0.14 & -0.40 \\ -0.76 & -0.40 & 7.91 \end{bmatrix}. \quad (10.20)$$

10.5 Summary and outlook

In theory, the ellipsoidal height h, measured by GPS, should be equal to the sum of the gravimetric geoid N and the orthometric height H measured by levelling campaigns. However, this is in general not valid due to systematic and stochastic errors among the three measurement types. A solution to this problem is to estimate a corrector surface and add this to the gravimetric geoid. In this way, we obtain an height reference surface, which can be used to transform ellipsoidal (GPS) heights to orthometric heights.

We have computed such a corrector surface for Switzerland. The three data sets (GPS, levelling and gravimetric geoid) were provided with full variance-covariance information. Comparisons have shown that we can not neglect the covariances. This will

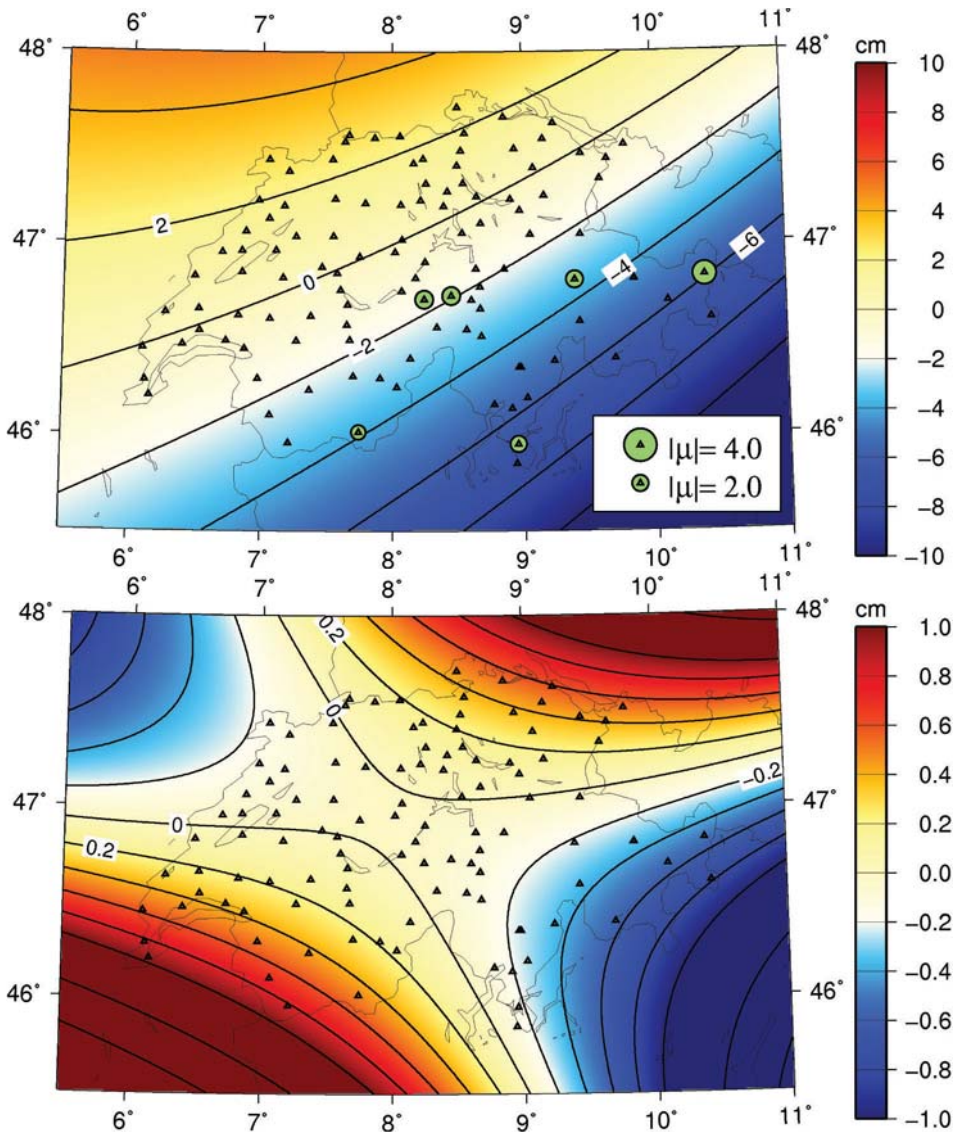


Fig. 10.6: Top figure: chosen corrector surface ($f_1'(\phi, \lambda)$), with the locations of the down-weighted observations. The bottom figure shows the differences with the corrector surface, using the four parameter form.

disturb the estimation of the corrector surface, the variance components, and the outlier detection.

Improvements of the corrector surface were possible with a combination of VCE (MINQUE) and a robust M-estimation technique (Huber). In this way, several GPS-levelling points had to be down-weighted. This was done in an iterative process with the re-weighting of the cofactor matrices, in which the cofactor matrix of the levelling data had to be rescaled by a factor 4. The tests were performed using a conventional four parameter form as parametrization of the corrector surface.

We have compared several parametrizations of the corrector surface and tested the parameters for significance. The best parametrization, using only 3 parameters, showed differences in the corrector surface with the conventional four parameter form up to 0.5 cm.

11

Conclusions and recommendations

11.1 Conclusions

Since the beginning of the space era in 1957 (with the launch of the Sputnik satellite), satellite measurements have been used in the estimation of the gravity field of the Earth. The combination of different data sets is essential in this estimation. It can fill the gaps in the space domain (e.g., polar gaps), strengthen parts of the frequency domain or extend the time span of the observations. Such combinations can include different arcs within the same satellite mission, arcs of different satellite missions, and the combination of satellite gravity measurements with other (prior) information, such as surface gravity data.

In this thesis, we propose to combine a validation of the stochastic model with a proper modelling of the functional model and a robust treatment of the outliers. As these algorithms are connected to each other, it is not possible to obtain an accurate estimate of the stochastic model without correctly augmenting the functional model or without a proper treatment of the outliers. The weighted least-squares solution should therefore be an iterative process.

The algorithms developed in this thesis are applied to a combination of CHAMP pseudo-observations with a prior gravity model, to a combination of GRACE monthly gravity solutions with weekly GPS site displacements, to the reduction of the temporal aliasing effect, and to the estimation of a corrector surface for Switzerland using a combination of levelling data, ellipsoidal heights (GPS) and a gravimetric geoid.

The functional model describes the linear relationship between the observations and the common (global) parameters we are interested in, e.g., the potential coefficients of a global gravity model. Moreover, it should account for systematic effects in the data and other inconsistencies with the other data sets. This can be done by an augmentation

of the functional model, i.e., the addition of (local) parameters to the functional model to account for these effects. In this way, we could estimate weekly estimates of the motion of the geocentre in the GPS / GRACE combination, which could be considered as local parameters of the GPS site displacements. However, the addition of the local parameters should first be tested for significance. This is done by hypothesis testing, i.e. we compare the estimates of the parameters with their quality description. As the latter depends on the stochastic model of the observations, we need to have a proper stochastic model in order to test the local parameters of the functional model. We have conducted such a testing on inconsistencies between the CHAMP-only solution and the prior gravity model, on inconsistencies between the GPS site displacements and the GRACE solution, and on the different parametrizations of the corrector surface for Switzerland.

The stochastic model describes the noise behaviour of the observations and is needed in the weighted least-squares estimation, the testing of the local parameters and the detection and treatment of the outliers. Moreover, it propagates into the quality description of the vector of unknowns and consequently the geophysical by-products. The noise level (and covariance structure) of the observations can change in time, due to, e.g., instrument conditions or a changing GPS constellation. However, if signal is not properly taken care of in the functional model (e.g., systematic effects, truncation of the global models, and temporal aliasing effects), this signal will be absorbed by the vector of residuals and should be accounted for in the stochastic model. We have tried to reduce the influence of the temporal aliasing of daily variations in the land hydrology on the monthly solutions by a refinement of the stochastic model. This was possible in a noise-free simulation. However, in the present test setup, the noise level was much higher than the aliasing signal. In this way, the validation of the stochastic model mainly followed the noise variances of the simulated data set.

To improve the stochastic model, we first have to write the variance-covariance (VC) matrix of the observations as a linear combination of several cofactor matrices. The estimation of these linear coefficients (variance components) is called variance component estimation (VCE). The most commonly used VCE method is the Minimum Norm Quadratic Unbiased Estimator (MINQUE), which, under normality, converges to (restricted) maximum likelihood, has minimum variance and can be derived in a least-squares approach. The latter property enables us to test the variance components for significance, add prior information to the variance components and linearize the VCE in non-linear expressions of the VC-matrix. In this thesis, we could for the first time derive the equations for MINQUE under the assumption of uncorrelated observation groups and with multiple variance components to be estimated for each observation group.

A comparison is made between the unbiased estimates of MINQUE and the alternative methods of the Iterative Restricted Maximum Likelihood Estimator (IREML), Helmert's VCE, Lerch's subset solution method and the Iterative Maximum Likelihood Estimator (IMLE). Helmert's VCE is identical to MINQUE if and only if we assume uncorrelated observation groups. The other three estimators are biased at each iteration, but converge to MINQUE for large time series of satellite gravity data, as was shown in

the gravity recovery using the CHAMP pseudo-observations. In small-scale problems, e.g., network adjustments, one should only use MINQUE or IREML, as they produce unbiased estimates at convergence.

The number of available satellite gravity measurements has increased throughout the years. Large improvements in the estimation of the global gravity field were obtained with the launch of the GEOS-3 satellite in 1975, the LAGEOS-I satellite in 1976, the TOPEX / Poseidon satellite in 1992, and recently with the launch of the CHAMP satellite in 2000 and the GRACE satellite mission in 2002. The upcoming GOCE satellite will further improve the estimation of the static gravity field of the Earth.

We can see a gradual change in the data weighting methods throughout the years, starting with trial and error methods, then external calibrations, by comparing the results to independent data, and finally internal calibration methods (such as Lerch's subset solution method and variance component estimation). However, with the launch of the dedicated satellite missions, most institutes have used a simple equal weighting approach to weight the different data sets, probably due to the elaborate processing of large amounts of data. We have derived Monte Carlo algorithms, which enable us to use existing least-squares software as a black box to obtain the variance components. In this way, the refinement of the stochastic model (and consequently an improvement of the least-squares estimate) hardly needs any extra computation time. One can simply use their own software package, change the observation vector into a randomized vector and use the output to derive the variance components. In this study, we could for the first time derive the equations of the Monte Carlo variant of the Minimum Norm Quadratic Unbiased Estimator (MINQUE).

We have compared the Monte Carlo variants of the several estimators (MINQUE, IREML, IMLE and Lerch) in the gravity recovery using the CHAMP pseudo-observations. The computational costs involved in the computation of MC-MINQUE and MC-Lerch is several orders larger than the computational costs involved in the estimation of MC-IREML and IMLE. We can therefore conclude that the MC-IREML is a fast, unbiased method, which can be used in satellite gravity field modelling using large amounts of data.

When we estimate the vector of unknowns or the variance components, we assume that outliers are treated in a correct way. The treatment of the outliers starts with a proper detection of the outliers. One should be cautious to assume correlations among the observations in the detection of the outliers, as outliers in general do not match the assumed functional and stochastic model. Moreover, large number of outliers may occur, which do not correspond to the stochastic model. We therefore choose to neglect the covariances in the detection of the outliers in large time series of observations. These covariance are however not neglected in the weighted least-squares estimation of the vector of unknowns or in the VCE. In the inversion of GPS site displacements into surface mass anomalies, we could see that neglecting the covariances would lead to errors, which could range up to half of the amplitude of the surface mass signal. Moreover, it was shown that neglecting the covariances in the estimation of the corrector surface of Switzerland had a significant effect on the estimation of the variance components.

We have compared robust M-estimation techniques to the conventional 3-sigma removal technique to treat the outliers in a data set. In a M-estimation, one assumes a higher probability for outliers and in this way down-weight the outliers. To do this, one needs to define an alternative distribution to the normal distribution, e.g., the Huber distribution, which is normally distributed within a certain interval and follows a Laplace distribution outside this interval. We have further developed the Cost Function Estimation (CFE). With this method, we no longer pre-define the distribution of the observations, but estimate this distribution from the data itself. In this way, not only observations with a high test statistic are re-weighted, but the re-weighting involves all observations, as these data are in general not normally distributed. A simulation study and the gravity recovery of real CHAMP pseudo-observations showed that the CFE is a proper robust weighting algorithm, which performs better than any other outlier treatment technique. Moreover, the determination of the cost function is very fast, as it makes use of the standardized residuals. A disadvantage is that all elements of the normal matrix need to be re-weighted.

The influence of the outliers on the estimation of the variance components should also be reduced. We can use the robust weights from the M-estimation or the CFE to down-weight the residuals of the outliers in the VCE algorithm. In this way, the re-weighting in the estimation of the vector of unknowns and the variance components is consistent, contrary to Fellner's approach, which is often used in robust variance component estimation. One should be cautious for descending influence functions, which can arise in the M-estimation or the CFE, as they can produce too optimistic VC-matrices. A solution to this problem is to use the Huber distribution in the robust VCE.

11.2 Recommendations

In this section, some recommendations for further research will be addressed:

- One should further look into the problem of correlations among the observations in the detection of outliers. The covariances in the stochastic model will smooth outliers over multiple residuals. The assumption of these correlations will de-smooth these residuals. In this way, the outliers should become more visible. However, this assumes that, after the adjustment of a single outlier, the other observations follow the stochastic model, which is in general not the case. Correlations may have changed after the blunder has occurred.

- The M-estimation method, and consequently the cost function estimation (CFE), assume the data to be uncorrelated. One should generalize the equations to account for these correlations.

- The cost function estimation may be improved by a different parametrization than the second order polynomials used so far.

- One could derive the REML-fs of the variance components if the distribution of the observations is estimated by a modified CFE, in which the expression of the function is not dependent on the standardized residuals.

- Future work should include covariances among the CHAMP pseudo-observations, as they are correlated due to the numerical Lagrange interpolator. This may have a positive effect on the rescaling of the EGM96 EVD cofactor matrices in the combination with the CHAMP-only solution.

- The algorithms, proposed in this thesis, should be applied to the temporal gravity field modelling using GRACE observations. In this way, outliers will be down-weighted in the data and a proper VC-matrix of the potential coefficients is obtained.

- The combination of weekly GPS site displacements with monthly GRACE models could be further improved by the inclusion of satellite laser ranging (low-degree coefficients) and the inclusion of ocean bottom pressure data to improve the oceanic regions. Moreover, the monthly solutions should be connected with each other by using an singular value decomposition in time and space or by computing one model, which accounts for the static part of the gravitational field and the main periodic and secular effects among the potential coefficients.

- The possible reduction of the aliasing effect by VCE should be further addressed by adding more signal to the simulated data set, e.g., ocean tide errors, atmospheric model errors and more hydrological signal.

- The upcoming GOCE mission deserves a proper modelling of the functional and stochastic models. Possible systematic effects should be tested for significance and augmented for in the functional model. The different sources of errors have to be defined and the variance-covariance matrix has to be written as a linear combination of cofactor matrices. Monte Carlo VCE is able to handle the large amount of data and the dimension of the problem. The standardized residuals can then be used to estimate the probability density function out of the data and consequently re-weight the observations. Combinations with other satellite missions (e.g., CHAMP and GRACE) are necessary, as well as with terrestrial, marine and airborne data to strengthen parts of the frequency domain, fill in the polar gaps and extend the time interval of the observations. A re-weighting of the different data sets using VCE will most likely improve the combined solution.

References

Amiri-Simkooei, A.R. (2007). *Least-Squares Variance Component Estimation*, PhD thesis, Delft University of Technology, Publications on Geodesy, 64, Netherlands Geodetic Commission, Delft

Amiri-Simkooei, A.R., C.C.J.M. Tiberius, P.J.G. Teunissen (2007). *Assessment of noise in GPS coordinate time series: Methodology and results*, Journal of Geophysical Research, vol. 112, B07413, doi: 10.1029/2006JB004913

Baarda, W. (1968). *A testing procedure for use in geodetic networks*, Publications on Geodesy, Vol. 2, No. 5, Netherlands Geodetic Commission, Delft

Balmino, G., B. Moynet, C. Reigber (1976a). *The GRIM2 earth gravity field model*, Reihe A No. 086, Deutsche Geodätische Kommission, Munich

Balmino, G., C. Reigber, B. Moynet (1976b). *A geopotential model determined from recent satellite observing campaigns (GRIM1)*, Manuscripta Geodaetica 1 (1), p. 41-69

Beaton, A.E. and J.W. Tukey (1974). *The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data*, Technometrics, vol. 16, p. 147-185

Biancale, R., G. Balmino, J.-M. Lemoine, J.-C. Marty, B. Moynet, F. Barlier, P. Exertier, O. Laurain, P. Gegout, P. Schwintzer, C. Reigber, A. Bode, R. König, F.-H. Massmann, J.-C. Raimondo, R. Schmidt, S.Y. Zhu (2000). *A new global Earth's gravity field model from satellite orbit perturbations: GRIM5-S1*, Geophysical Research Letters 27 (22), p. 3611-3614

Blewitt, G. (2003). *Self-consistency in reference frames, geocenter definition, and surface loading of the solid Earth*, Journal of Geophysical Research, vol. 108, no. B2, 2103, doi: 10.1029/2002JB002082

Blewitt, G. and P. Clarke (2003). *Inversion of Earth's changing shape to weigh sea level in static equilibrium with surface mass redistribution*, Journal of Geophysical Research, 108, B6, 2311, doi: 10.1029/2002JB002290

Blewitt, G., D. Lavallee, P. Clarke, K. Nurutdinov (2001). *A new global mode of Earth deformation: Seasonal cycles detected*, Science, 294, p. 2342-2345

Bouman, J. (1998a). *Quality of regularization methods*, DEOS Report no. 98.2, Delft University Press, Delft

Bouman, J. (1998b). *A survey of global gravity models*, DEOS Report no. 98.3, Delft University Press, Delft

Buchar, E. (1958). *Motion of the nodal line of the second Russian Earth satellite (1957 β) and flattening of the Earth*, Nature, 182, p. 198-199

Chang, X.-W. and Y. Guo (2005). *Huber's M-estimation in relative GPS positioning: computational aspects*, Journal of Geodesy, 79, p. 351-362

Chao, B.F. (2005). *On inversion for mass distribution from global (time-variable) gravity field*, Journal of Geodynamics, 39, p. 223-230

Chao, B.F. and A.Y. Au (1991). *Temporal variations of the Earth's low-degree zonal gravitational field caused by atmospheric mass redistribution: 1980-1988*, Journal of Geophysical Research, 96, B4, p. 6,569-6,575

Cheng, J.L., C.R. Wilson, K.-W. Seo (2006). *Optimized smoothing of Gravity Recovery and Climate Experiment (GRACE) time-variable*, Journal of Geophysical Research, 111, B06408, doi:10.1029/2005JB004064

Constable, C.G. (1988). *Parameter estimation in non-Gaussian noise*, Geophysical Journal, 94, p. 131-142

Constable, C.G. and R.L. Parker (1988). *Smoothing, splines and smoothing splines; their application in geomagnetism*, Journal of Computational Physics, 78, p. 493-508

Cramér, H. (1946). *Mathematical Methods of Statistics*, Princeton University Press, Princeton

Crocetto, N., M. Gatti. P. Russo (2000). *Simplified formulae for the BIQUE estimation of variance components in disjunctive observation groups*, Journal of Geodesy, 74, p. 447-457

CSR (2008). *GRACE, Gravity Recovery and Climate Experiment*, http://www.csr.utexas.edu/grace, CSR, University of Texas

De Min, E.J. (1996). *De geoïde voor Nederland*, PhD thesis (in Dutch), Delft University of Technology, Delft

Ding, X. and R. Coleman (1996). *Multiple outlier detection by evaluating redundancy contributions of observations*, Journal of Geodesy, 70, p. 489-498

Ditmar, P. and A.A. van Eck van der Sluijs (2004). *A technique for modeling the Earth's gravity field on the basis of satellite accelerations*, Journal of Geodesy, 78, p. 12-33

Ditmar, P., V. Kuznetsov, A.A. van Eck van der Sluijs, E. Schrama, R. Klees (2006). *'DEOS_CHAMP-01C_70':a model of the Earth's gravity field computed from accelerations of the CHAMP satellite*, Journal of Geodesy, 79, p. 586-601

Ditmar, P., R. Klees, X. Liu (2007). *Frequency-dependent data weighting in global gravity field modeling from satellite data contaminated by non-stationary noise*, Journal of Geodesy, 81, p. 81-96

Dong, D., T. Yunck, M. Heflin (2003). *Origin of the International Terrestrial Reference Frame*, Journal of Geophysical Research, vol. 108, no. B4, 2200

Drinkwater, M.R., R. Haagmans, D. Muzi, A. Popescu, R. Floberghagen, M. Kern, M. Fehringer (2007). *The GOCE Gravity Mission: ESA's First Core Earth Explorer*, Proceedings of the 3rd International GOCE User Workshop, 6-8 November, 2006, Frascati, Italy, ESA SP-627, p. 1-8

Dueck, A. and S. Lohr (2005). *Robust estimation of multivariate covariance components*, Biometrics, 61, p. 162-169

Egbert, G.D. and S.Y. Erofeeva (2002). *Efficient Inverse modeling of barotropic ocean tides*, Journal of Atmospheric and Oceanic Technology, vol. 19, 2, p. 183-204

Egeltoft, T. (1992). *Variance component estimation in geodetic networks*, TRITA/ GEOD report no. 92/1027, Royal Institute of Technology, Department of Geodesy, Stockholm

Farrell, W.E. (1972). *Deformation of the Earth by Surface Loads*, Reviews of Geophysics and Space Physics, 10, 3, p. 761-797

Featherstone, W. (1998). *Do we need a gravimetric geoid or a model of the Australian Height Datum to transform GPS heights in Australia?*, Australian Surveyor, 43, 4, p. 273-280

Fellner, W.H. (1986). *Robust estimation of variance components*, Technometrics, vol. 28, no. 1, p. 51-60

Ferland, R., J. Kouba, D. Hutchison (2000). *Analysis methodology and recent results of the IGS network combination*, Earth Planets Space, 52, p. 953-957

Földvary, L. D. Švehla, C. Gerlach, M. Wermuth, T. Gruber, R. Rummel, M. Rothacher, B. Frommknecht, T. Peters, P. Steigenberger (2005). *Gravity model TUM-2Sp based on the energy balance approach and kinematic CHAMP orbits*. In: Reigber, C., H. Lühr, P. Schwintzer, J. Wickert (eds.), *Earth Observation with CHAMP, Results from Three Years in Orbit*, p. 13-18, Springer

Förste, C., R. Schmidt, R. Stubenvoll, F. Flechtner, U. Meyer, R. K´onig, H. Neumayer, R. Biancale, J.-M. Lemoine, S. Bruinsma, S. Loyer, F. Barthelmes, S. Esselborn (2008). *The GeoForschungsZentrum Potsdam / Groupe de Recherche de Géodèsie Spatiale satellite-only and combined gravity field models: EIGEN-GL04S1 and EIGEN-GL04C*, Journal of Geodesy, 82, p. 331-346

Förstner, W. (1979). *Ein Verfahren zur Schätzung von Varianz-Kovarianz Komponenten.*, Allgemeine Vermessungs-Nachrichten, 86, p. 446-453

Fotopoulos, G. (2003). *An analysis on the optimal combination of geoid, orhtometric and ellipsoidal height data*, PhD thesis, UCGE Reports no. 20185, Department of Geomatics Engineering, Univserity of Calgary

Fotopoulos, G. (2005). *Calibration of geoid error models via a combined adjustment of ellipsoidal, orthometric and gravimetric geoid height data*, Journal of Geodesy, 79, p. 111-123

Galloway, D.L., K.W. Hudnut, S.E. Ingebritsen, S.P. Philips, G. Peltzer, F. Rogez, and P.A. Rosen (1998). *Detection of aquifer system compaction and land subsidence using interferometric synthetic aperture radar. Antelope Valley, Mojave Desert, California*, Water Resources Research, 34, 10, p. 2573-2585

Gaposchkin, E.M. (1973). *1973 Smithsonian Standard Earth (III)*, Smithsonian Institution Astrophyscial Observatory, Special report no. 353

Gerlach, C. L. Földvary, D. Švehla, Th. Gruber, M. Wermuth, N. Sneeuw, B. Frommknecht, H. Oberndorfer, Th. Peters, M. Rothacher, R. Rummel, P. Steigenberger (2003). *A CHAMP-only gravity field model from kinematic orbits using the energy balance integral*, Geophysical Research Letters, 30, 20, 2037

GFZ (2003). *GRACE gravity model EIGEN-GRACE01S*, http://www.gfz-potsdam.de/grace/results/grav/g001_EIGEN_GRACE01S.pdf

Girard, D.A. (1989). *A fast 'Monte Carlo Cross-Validation' procedure for large least-squares problems with noisy data*, Numerische Mathematik 56, p. 1-23

Girard, D.A. (1995). *The fast Monte Carlo cross-validation and C_l procedures: Comments, new results and application to image recovery problems*, Computational Statistics, 10, p. 205-231

Golub, G.H. and U. von Matt (1997). *Generalized cross-validation for large-scale problems*, Journal of Computational and Graphical Statistics, 6, p. 1-34

Golub, G.H., M. Heath, G. Wahba (1979). *Generalized cross-validation as a Method for Choosing a Good Ridge Parameter*, Technometrics, vol. 21, no. 2, p. 215-223

Götzelmann, M., W. Keller, T. Reubelt (2006). *Gross error compensation for gravity field analysis based on kinematic orbit data*, Journal of Geodesy, 80, p. 184-198

Grafarend, E., A. Kleusberg, B. Schaffrin (1980). *An introduction to the Variance-Covariance- Component Estimation of Helmert type*, Zeitschr. f. Verm., 4, p. 161-180

Grodecki, J. (1999). *Generalized maximum-likelihood estimation of variance components with inverted gamma prior*, Journal of Geodesy, 73, p. 367-374

Gruber, T., A. Bode, C. Reigber, P. Schwintzer, G. Balmino, R. Biancale, J.-M. Lemoine (2000). *GRIM5-C1: Combination solution of the global gravity field*, Geophysical Research Letters 27 (24), p. 4005-4008

Han, S.-C. (2004). *Efficient determination of global gravity field from satellite-to-satellite tracking mission*, Celestial Mechanics and Dynamical Astronomy, 88, p. 69-102

Han, S.-C. and F.J. Simons (2008). *Spatiospectral localization of global geopotential fields from the Gravity Recovery and Climate Experiment (GRACE) reveals the coseismic gravity change owing to the 2004 Sumatra-Andaman earthquake*, Journal of Geophysical Research, vol. 113, B01405, doi: 10.1029/2007JB004927

Han, S.-C., C. Reigber, C.K. Shum (2002). *Efficient gravity field recovery using in situ disturbing potential observables from CHAMP*, Geophysical Research Letters, 29 (16), article 36

Han, S.-C., C. Jekeli, C.K. Shum (2004). *Time-variable aliasing effects of ocean tides, atmosphere, and continental water mass on monthly mean GRACE gravity field*, Journal of Geophysical Research, 109, B04403,
doi: 1029/2003JB002501

Han, S.-C., C.K. Shum, A. Braun (2005a). *High-resolution continental water storage recovery from low-low satellite-to-satellite tracking*, Journal of Geodynamics, 39, p. 11-28, doi: 10.1016/j.jog.2004.08.002

Han, S.-C., C.K. Shum, C. Jekeli, D. Alsdorf (2005b). *Improved estimation of terrestrial water storage changes from GRACE*, Geophysical Research Letters, 32, L07302, doi:10.1029/2005GL022382

Han, S.-C., C.K. Shum, C. Jekeli, C.-Y. Kuo, C. Wilson (2005c). *Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement*, Geophysical Journal International, 163, p. 18-25,
doi: 10.1111/j.1365-246X.2005.02756.x

Han, S.-C., C.K. Shum, P. Ditmar, P. Visser, C. van Beelen, E.J.O. Schrama (2006). *Aliasing effect of high-frequency mass variations on GOCE recovery of the earth's gravity field*, Journal of Geodynamics, 41, p. 69-76

Hardy, R.L. (1971). *Multiquadric equations of topography and other irregular surfaces*, Journal of Geophysical Research, 76, 8, p. 1905-1915

Harville, D.A. (1977). *Maximum likelihood approaches to variance component estimation and to related problems*, Journal of the American Statistical Association, 72, p. 320 - 338

Heiskanen, W.A. and H. Moritz (1967). *Physical Geodesy*, W.H. Freeman and Company, San Fransisco (reprint 1996, Institute of Physical Geodesy, Technical University, Graz)

Helmert, F.R. (1924). *Die Ausgleichsrechnung nach der Methode der kleinste Quadrate*, 3. Aufl., Leipzig-Berlin

Horn, S.D., R.A. Horn, D.B. Duncan (1975). *Estimating heteroscedastic variances in linear models*, Journal of the American Statistical Association, 70, p. 380-385

Howe, E. and C. C. Tscherning (2003). *Preliminary analysis of CHAMP state vector and accelerometer data for the recovery of the gravity potential.*. In: Reigber, C., H. Lühr, P. Schwintzer (ed.). *First CHAMP Mission results for Gravity, Magnetic and Atmospheric Studies*, Springer, p. 140-145

Huber, P.J. (1964). *Robust estimation of a location parameter*, Annals of Mathematical statistics, 35, p. 73-101

Huber, P.J. (1981). *Robust Statistics*, Wiley series in probability and mathematical statistics, New York

Hutchinson, M.F. (1990). *A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines*, Communications in Statistics, Simulation and Computation 19, p. 43-450

Ilk, K.H., T. Mayer-Gürr, M. Feuchtinger (2005). *Gravity field recovery by analysis of short arcs of CHAMP*. In: Reigber, C., H. Lühr, P. Schwintzer, J. Wickert (eds.), *Earth Observation with CHAMP, Results from Three Years in Orbit*, p. 127-132

Jansen, M.J.F., B.C. Gunter, J. Kusche (2008a). *Evaluating the combination of grace, gps and ocean bottom pressure data on global mass transport estimates*, submitted to the Geophysical Journal International

Jansen, M.J.F., B.C. Gunter, R. Rietbroek, C. Dahle, J. Kusche, F. Flechtner, S.-E. Brunnabend, J. Schröter (2008b). *Estimating sub-monthly global mass transport signals using GRACE, GPS and OBP data sets*, submitted to the proceedings

of the IAG International Symposium on Gravity, Geoid and Earth Observation, Chania, Crete, 23-27 June, 2008

Jekeli, C. (1981). *Alternative methods to smooth the Earth's gravity field*, OSU report no. 327, Ohio State University, Columbus

Jekeli, C. (1999). *The determination of gravitational potential differences from satellite-to-satellite tracking*, Celestial Mechanics and Dynamical Astronomy, 75, p. 85-101

Jekeli, C. (2001). *On the determination of geopotential differences from satellite-to-satellite tracking*. In: Sideris, M.G. (Ed.), *IV Hotine-Marussi Symposium on Mathematical Geodesy*, Springer, p. 33-39

Jiang, Z. and H. Duquenne (1996). *On the combined adjustement of a gravimetrically determined geoid and GPS levelling stations*, Journal of Geodesy, 70, p. 505-514

Kang, Z., S. Bettadpur, B. Tapley, M. Cheng, J. Ries (2003). *Determination of CHAMP accelerometer calibration parameters*. In: Reigber, C., H. Lühr, P. Schwintzer (ed.). *First CHAMP Mission results for Gravity, Magnetic and Atmospheric Studies*, Springer, p. 19-25

Kaula, W.M. (1961). *A geoid and world geodetic system based on a combination of gravimetric, astro-geodetic, and satellite data*, Journal of Geophysical Research, 66, 6, p. 1799-1811

Kaula, W.M. (1966). *Theory of Satellite Geodesy*, Blaisdell Publ. Company, Waltham, Mass., (reprint 2000, Dover Publ., New York)

Kelm, R. (1978). *Ist die Varianzschätzung nach Helmert MINQUE?*, Allg. Verm. nachr., 85, p. 49-54

Kern, M., T. Preimesberger, M. Allesch, R. Pail, J. Bouman, R. Koop (2005). *Outlier detection algorithms and their performances in GOCE gravity field processing*, Journal of Geodesy, 78, p. 509-519

Kizilsu, G., M. Sahin (2000). *SLR precision analysis for LAGEOS I and II*, Earth Planets Space, 52, p. 789-794

Koch, K.R. (1986). *Maximum likelihood estimate of variance components*, Bull. Géod., 60, p. 329-338

Koch, K.R. (1990). *Bayesian Inference with Geodetic Applications*, Lecture Notes in Earth Sciences, Springer

Koch, K.R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models*, second edition, Springer

Koch, K.R. and J. Kusche (2002). *Regularization of geopotential determination from satellite data by variance components*, Journal of Geodesy, 76, p. 259-268

Koch, K.R. and F. Morrison (1970). *A simple layer model of the geopotential from a combination of satellite and gravity data*, Journal of Geophysical Research, 75, 8, p. 1483-1492

Kotsakis, C. (2005). *A type of biased estimators for linear models with uniformly biased data*, Journal of Geodesy, 79, p. 341-350

Kotsakis, C. and M.G. Sideris (1999). *On the adjustment of combined GPS/ levelling/ geoid networks*, Journal of Geodesy, 73, p. 412-421

Kusche, J. (2003). *A Monte-Carlo technique for weight estimation in satellite geodesy*, Journal of Geodesy, 76, p. 641-652

Kusche, J. (2007). *Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models*, Journal of Geodesy, 81, p. 733-749, doi:10.1007/s00190-007-0143-3

Kusche, J. and R. Klees (2002). *Regularization of gravity field estimation from satellite gravity gradients*, Journal of Geodesy, 76, p. 359-368

Kusche, J. and E.J.O. Schrama (2005). *Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data*, Journal of Geophysical Research, 110, B09409, doi:10.1029/2004JB003556

Kusche, J. and J.P. van Loon (2005). *Statistical assessment of CHAMP data and models using the Energy Balance Approach*. In: Reigber, C., H. Lühr, P. Schwintzer, J. Wickert (eds.), *Earth Observation with CHAMP, Results from Three Years in Orbit*, p. 133-138, Springer

Kusche, J., E.J.O. Schrama, M.J.F. Jansen (2007). *Continental hydrology retrieval from GPS time series and GRACE gravity solutions*. In: Tregoning, P. and C. Rizos (eds.). *Dynamic Planet, Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools*, IAG Symposia, vol. 130, Springer, p. 517-522

Lambeck, K. and R. Coleman (1983). *The Earth's shape and gravity field: a report of progress from 1958 to 1982*, Geophysical Journal of the Royal Astronomical Society, 74, p. 25-54

Le Grand, P. and J.F. Minster (1999). *Impact of the GOCE gravity mission on ocean circulation estimates*, Geophysical Research Letters, 26 (13), p. 1881-1884

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

Lerch, F.J. (1989). *Optimum data weighting and error calibration for estimation of gravitational parameters*, NASA Technical Memorandum 100737, NASA, Goddard Space Flight Center

Lerch, F.J. (1991). *Optimum data weighting and error calibration for estimation of gravitational parameters*, Bulletin Géodésique, 65, p. 44-52

Lerch, F.J., S.M. Klosko, R.E. Laubscher, C.A. Wagner (1979). *Gravity model improvement using Geos-3 (GEM 9 and 10)*, Journal of Geophyscial Research, 95, B8, p. 3897-3916

Lerch, F.J., B.H. Putney, C.A. Wagner, S.M. Klosko (1981). *Goddard earth models for oceanographic applications (GEM 10B and 10C)*, Marine Geodesy, 5(2), p. 145-187

Lerch, F.J., S.M. Klosko, G.B. Patel (1982). *A refined graivty model from LAGEOS (GEM-L2)*, Geophysical Research Letters, vol. 9, no. 11, p. 1263-1266

Lerch, F.J., S.M. Klosko, G.B. Patel, C.A. Wagner (1985). *A gravity model for crustal dynamics (GEM-L2)*, Journal of Geophysical Research, 90, B11, p. 9301-9334

Lerch, F.J., J.G. Marsh, S.M. Klosko, E.C. Pavlis, G.B. Patel, D.S. Chinn, C.A. Wagner (1988). *An improved error assessment for the GEM-T1 gravitational model*, NASA Tecnical Memorandum 100713, NASA, Goddard Space Flight Center

Lerch, F.J., J.G. Marsh, S.M. Klosko, G.B. Patel, D.S. Chinn, E.C. Pavlis, C.A. Wagner (1991). *An improved error assessment for the GEM-T1 gravitational model*, Journal of Geophysical Research, 96, p. 20023-20040

Lerch, F.J., R.S. Nerem, D.S. Chinn, J.C. Chan, G.B. Patel, S.M. Klosko (1993). *New error calibration tests for gravity models using subset solutions and independent data: applied to GEM-T3*, Geophysical Research Letters, vol. 20, no. 2, p. 249-252

Lerch, F.J., R.S. Nerem, B.H. Putney, T.L. Felsentreger, B.V. Sanchez, J.A. Marshall, S.M. Klosko, G.B. Patel, R.G. Williamson, D.S. Chinn, J.C. Chan, K.E. Rachlin, N.L. Chandler, J.J. McCarthy, S.B. Luthcke, N.K. Pavlis, D.E. Pavlis, J.W. Robbins, S. Kapoor, E.C. Pavlis (1994). *A geopotential model from satellite tracking, altimeter, and surface gravity data: GEM-T3*, Journal of Geophysical Research, 99, B2, p. 2815-2839

Li, X. and H.-J. Götze (2001). *Tutorial Ellipsoid, geoid, gravity, geodesy, and geophysics*, Geophysics, vol. 66, no. 6, p. 1660-1668

Löcher, A. and K.H. Ilk (2007). *A validation procedure for satellite orbits and force function models based on a new balance equation approach*. In: Tregoning, P. and C. Rizos (eds.). *Dynamic Planet, Monitoring and Understanding a Dynamic Planet with Geodetic And Oceanographic Tools*, IAG Symposia, vol. 130, Springer, p. 280-287

Lucas, J.R. (1985). *A variance component estimation method for sparse matrix applications*, NOAA Technical Report NOS 111 NGS 33, National Geodetic Survey, Rockville

Lucas, J.R. and W.H. Dillinger (1998). *MINQUE for block diagonal bordered systems such as encountered in VLBI data analysis*, Journal of Geodesy, 72, p. 343-349

Luthcke, S.B., H.J. Zwally, W. Abdalati, D.D. Rowlands, R.D. Ray, R.S. Nerem, F.G. Lemoine, J.J. McCarthy, D.S. Chinn (2006). *Recent Greenland ice mass loss by drainage system from satellite gravity observations*, Science, vol. 314, p. 1286-1289

Madsen, K. and H.B. Nielsen (1990). *Finite algorithms for robust linear regression*, BIT, 30, p. 682-699

Marsh, J.G., F.J. Lerch, B.H. Putney, D.C. Christodoulidis, T.L. Felsentreger, B.V. Sanchez, D.E. Smith, S.M. Klosko, T.V. Martin, E.C. Pavlis, J.W. Robbins, R.G. Williamson, O.L. Colombo, N.L. Chandler, K.E. Rachlin, G.B. Patel, S. Bhati, D.S. Chinn (1987). *An improved model of the Earth's gravitational field: GEM-T1*, NASA Technical Memorandum 4019, NASA, Goddard Space Flight Center

Marsh, J.G., F.J. Lerch, B.H. Putney, D.C. Christodoulidis, D.E. Smith, T.L. Felsentreger, B.V. Sanchez, S.M. Klosko, E.C. Pavlis, T.V. Martin, J.W. Robbins, R.G. Williamson, O.L. Colombo, D.D. Rowlands, W.F. Eddy, N.L. Chandler, K.E. Rachlin, G.B. Patel, S. Bhati, D.S. Chinn (1988). *A new gravitational model for the Earth from satellite tracking data: GEM-T1*, Journal of Geophysical Research, 93, B6, p. 6169-6215

Marsh, J.G., F.J. Lerch, B.H. Putney, T.L. Felsentreger, B.V. Sanschez, D.E. Smith, S.M. Klosko, E.C. Pavlis, J.W. Robbins, R.G. Williamson, O.L. Colombo, D.D. Rowlands, W.F. Eddy, N.L. Chandler, K.E. Rachlin, G.B. Patel, S. Bhati, D.S. Chinn (1989a). *Gravitational model improvement at the Goddard Space Flight Center*, in Rapp, R.H. (1989). *Progress in the determination of the Earth's gravity field*, OSU Report no. 397, Ohio State University, Columbus

Marsh, J.G., F.J. Lerch, B.H. Putney, T.L. Felsentreger, B.V. Sanchez, S.M. Klosko, G.B. Patel, J.W. Robbins, R.G. Williamson, T.E. Engelis, W.F. Eddy, N.L. Chandler, D.S. Chinn, S. Kapoor, K.E. Rachlin, L.E. Braatz, E.C. Pavlis (1989b). *The GEM-T2 Gravitational Model*, NASA Technical Memorandum 100746, NASA, Goddard Space Flight Center

Marsh, J.G., F.J. Lerch, B.H. Putney, T.L. Felsentreger, B.V. Sanchez, S.M. Klosko, G.B. Patel, J.W. Robbins, R.G. Williamson, T.L. Engelis, W.F. Eddy, N.L. Chandler, D.S. Chinn, S. Kapoor, K.E. Rachlin, L.E. Braatz, E.C. Pavlis (1990). *The GEM-T2 Gravitational Model*, Journal of Geophysical Research, 95, B13, p. 22043-22071

Marti, U. (2007). *Comparison of high precision geoid models in Switzerland.* In: Tregoning, P. and C. Rizos (eds.). *Dynamic Planet, Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools*, IAG Symposia, vol. 130, Springer, p. 377-382

Marti, U. and A. Schlatter (2005). *Festlegung des Höhenbezugsrahmens LHN95 und Berechnung des Geoidmodells CHGeo2004*, Geomatik Schweiz, 8/2005, p. 445 - 449

Marti, U., A. Schlatter, E. Brockmann, A. Wiget (2002). *The way to a consistent national height system in Switzerland.* In: Ádám, J. and K.-P. Schwarz (eds.). *Vistas for Geodesy in the New Millenium*, IAG Symposia, vol. 125, Sprinter, Berlin, p. 90-95

Matsumoto, K., T. Takanezawa, M. Ooe (2000). *Ocean tide models developed by assimilating TOPEX/Poseidon altimeter data into hydrodynamic model: a global model and a regional model around Japan*, J. Oceanogr., 56, p. 567-581

Mayer-Gürr, T., K.H. Ilk, A. Eicker, M. Feuchtinger (2005). *ITG-CHAMP01: a CHAMP gravity field model from short kinematic arcs over a one-year observation period*, Journal of Geodesy, 78, p. 462-480

Mayer-Gürr, T., A. Eicker, K.H. Ilk (2007). *ITG-GRACE02S: a GRACE gravity field derived from short arcs of the satellite orbit*. In: Harita Dergisi, proceedings of the 1st International Symposium of the International Gravity Field Service, Gravity Field of The Earth, 28 August - 1 September 2006, Istanbul, p. 193-198

Merson R.H. and D.G. King-Hele (1958). *Use of artificial satellites to explore the Earth's gravitational field: results from Sputnik 2 (1957β)*, Nature, 182, p. 640-641

Middel, B. and B. Schaffrin (1987). *Robust determination of the Earth's Gravity Potential Coefficients*, paper presented to the IAG Intersection Symp. *Advances in Gravity Field Modelling*, XIX. IUGG General Assembly, Vancouver

Milly, P.C.D. and A.B. Shmakin (2002). *Global modeling of land water and energy balances. Part I: The Land Dynamics (LaD) Model*, Journal of Hydrometeorology, 3, 3, p. 283-299

Mohr, P.J. and B.N. Taylor (2005). *CODATA recommended values of the fundamental physical constants: 2002*, Reviews of Modern Physics, vol. 77, p. 1-107

Molodenskii, M.S., V.F. Eremeev, M.I. Yurkina (1962). *Methods for study of the external gravitational field and figure of the Earth*, Israel Program for Scientific Translations, Jerusalem

Moritz, H. (1980). *Advanced Physical Geodesy*, Abacus, Tunbridge Wells

Nerem, R.S., F.J. Lerch, B.H. Putney, S.M. Klosko, G.B. Patel, R.G. Williamson, E.C. Pavlis (1992). *An improved model of the Earth's gravity field: GEM-T3*. In: Colombo O.L (ed.). *From Mars to Greenland: charting gravity with space and airborne instruments*, IAG symposia, vol. 110, Springer, p. 29-44

Nerem, R.S., B.H. Putney, J.A. Marshall, F.J. Lerch, E.C. Pavlis, S.M. Klosko, S.B. Luthcke, G.B. Patel, R.G. Williamson, N.P. Zelensky (1993). *Expected orbit determination performance for the TOPEX/Poseidon mission*, IEEE Transactions on Geoscience and Remote Sensing, vol. 31, no. 2, p. 333-353

Nerem, R.S., F.J. Lerch, R.G. Williamson, S.M. Klosko, J.W. Robbins, G.B. Patel (1994a). *Gravity model improvement using the DORIS tracking system on the SPOT 2 satellite*, Journal of Geophysical Research, 99, B2, p. 2791-2813

Nerem, R.S., F.J. Lerch, J.A. Marshall, E.C. Pavlis, B.H. Putney, B.D. Tapley, R.J. Eanes, J.C. Ries, B.E. Schutz, C.K. Shum, M.M. Watkins, S.M. Klosko, J.C. Chan, S.B. Luthcke, G.B. Patel, N.K. Pavlis, R.G. Williamson, R.H. Rapp, R. Biancala, F. Nouel (1994b). *Gravity model development for TOPEX/Poseidon: Joint Gravity Models 1 and 2*, Journal of Geophysical Research, 99, C12, p. 24421-24447

Nerem, R.S., C. Jekeli, W.M. Kaula (1995). *Gravity field determination and characteristics: Retrospective and prospective*, Journal of Geophysical Research, 100, B8, p. 15053-15074

Newton, I. (1687). *Philosophiae Naturalis Principia Mathematica*. Translated and published as: Cohen, B. and A. Whitman (1999). *The Principia: Mathematical Principles and Mathematical Philosophy*, University of California Press, Berkeley

O'Keefe, J.A. (1957). *An application of Jacobi's integral to the motion of an Earth satellite*, The Astronomical Journal, 62, 8, p. 265-266

Olsen, N. (2002). *A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data*, Geophysical Journal International, 149, p. 454-462

Ou, Z. (1991). *Approximative Bayes estimation for variance components*, Manuscripta Geodaetica, 16, p. 168-172

Ou, Z. (1993). *Bayesian inference for variance factor with maximum entropy prior*, Manuscripta Geodaetica, 18, p. 242-248

Ou, Z. and K.R. Koch (1994). *Analytical expressions for Bayes estimates of variance components*, Manuscripta Geodaetica, 19, p. 284-293

Patterson, H.D. and R. Thompson (1971). *Recovery of inter-block information when block sizes are unequal*, Biometrika, 58, 3, p. 545-554

Paul, M.K. (1983). *Recurrence relations for the truncation error coefficients for the extended Stokes function*, Bulletin Géodésique, 57, p. 152-166

Pavlis, N.K., S.A. Holmes, S.C. Kenyon, J.K. Factor (2008). *An Earth gravitational model to degree 2160: EGM2008*, oral presentation at the EGU General Assembly 2008, 13-18 April, 2008, Vienna.

Perosanz, F., R. Biancale, S. Loyer, J.-M. Lemoine, A. Perret, P. Touboul, B. Foulon, G. Pradels, L. Grunwald, Th. Fayard, N. Vales, M. Sarrailh (2003). *On Board Evaluation of the STAR Accelerometer*. In: Reigber, C., H. Lühr, P. Schwintzer (ed.). *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, Springer, p. 11-19

Pietrzak, J., A. Socquet, D. Ham, W. Simons, C. Vigny, R.J. Labeur, E. Schrama, G. Stelling, D. Vatvani (2007). *Defining the source region of the Indian Ocean tsunami from GPS, altimeters, tide gauges and tsunami models*, Earth and Planetary Science Letters, 261, p. 49-64

Pope, A.J. (1976). *The Statistics of Residuals and te Detection of Outliers*, NOAA Technical Report, NOS65, National Geodetic Information Center, NOS/NOAA, Rockville, Md.

Prutkin, I. and R. Klees (2008). *On the non-uniqueness of local quasi-geoids computed from terrestrial gravity anomalies*, Journal of Geodesy, 82, p. 147-156 , doi: 10.1007 / s00190-007-0161-1

Pukelsheim, F. (1976). *Estimating Variance Components in Linear Models*, Journal of Multivariate Analysis, 6, p. 626-629

Rao, C.R. (1971a). *Estimation of variance and covariance components - the MINQUE theory*, Journal of Multivariate Analysis, 1, p. 257-275

Rao, C.R. (1971b). *Minimum Variance Quadratic Unbiased Estimation of Variance Components*, Journal of Multivariate Analysis, 1, p. 445-456

Rao, C.R. (1973). *Linear statistical inference and its applications*, John Wiley and Sons, New York

Rao, C.R. and J. Kleffe (1988). *Estimation of variance components and applications*, Elsevier, Amsterdam

Rao, C.R. and S.K. Mitra (1971). *Generalized Inverse of Matrices and Its Applications*, John Wiley and Sons, New York

Rapp, R.H. (1968). *Gravitational potential of the Earth determined from a combination of satellite, observed, and model anomalies*, Journal of Geophysical Research, 73, 20, p. 6555-6562

Rapp, R.H. (1973). *Numerical results from the combination of gravimetric and satellite data using the principles of least squares collocation*, OSU Report no. 200, Ohio State University, Columbus

Rapp, R.H. (1978). *A global* $1^\circ \times 1^\circ$ *anomaly field combining satellite, Geos-3 altimeter and terrestrial data*, OSU Report no. 278, Ohio State University, Columbus

Rapp, R.H. (1981). *The Earth's gravity field to degree and order 180 using Seasat altimeter data, terrestrial gravity data, and other data*, OSU Report no. 322, Ohio State University, Columbus

Rapp, R.H. and J.Y. Cruz (1986a). *The representation of the Earth's gravitational in a spherical harmonic expansion to degree 250*, OSU Report no. 371, Ohio State University, Columbus

Rapp, R.H. and J.Y. Cruz (1986b). *Spherical harmonic expansions of the Earth's gravitational potential to degree 360 using* $30'$ *mean anomalies*, OSU Report no. 376, Ohio State University, Columbus

Rapp, R.H. and N.K. Pavlis (1990). *The development and analysis of geopotential coefficient models to spherical harmonic degree 360*, Journal of Geophysical Research 95, B13, p. 21885-21911

Rapp, R.H. and R. Rummel (1975). *Methods for the computation of detailed geoids and their accuracy*, OSU Report no. 233, Ohio State University, Columbus

Rapp, R.H., Y.M. Wang, N.K. Pavlis (1991). *The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models*, OSU Report no. 410, Ohio State University, Columbus

Ray, R.D. (1999). *A Global Ocean Tide Model From TOPEX/POSEIDON Altimetry: GOT99.2*, Report no. 209478, NASA, Greenbelt

Ray, R.D., D.D. Rowlands, G.D. Egbert (2003). *Tidal models in a new era of satellite gravimetry*, Space Science Reviews, 108, p. 271-282

Reigber, C. (1989). *Gravity field recovery from satellite data*. In: Sansò, F. and R. Rummel. *Theory of satellite geodesy and gravity field determination*. Lecture Notes in Earth Sciences, 25, Springer

Reigber, C., G. Balmino, B. Moynot, H. Müller (1983). *The GRIM3 earth gravity field model*, Manuscripta Geodaetica 8, p. 93-138

Reigber, C., P. Schwintzer, H. Lühr (1999). *The CHAMP geopotential mission*, Bollettino di Geofisica Teorica ed Applicata, Vol. 40, No. 3-4, p. 285-289

Reigber, C., G. Balmino, P. Schwintzer, R. Biancale, A. Bode, J.-M. Lemoine, R. König, S. Loyer, H. Neumayer, J.-C. Marty, F. Barthelmes, F. Perosanz, S. Y. Zhu (2002). *A high-quality global gravity model from CHAMP GPS tracking data and accelerometry (EIGEN-1S)*, Geophysical Research Letters 29 (14)

Reigber, C., P. Schwintzer, K.-H. Neumayer, F. Barthelmes, R. König, C. Förste, G. Balmino, R. Biancale, J.-M. Lemoine, S. Loyer, S. Bruinsma, F. Perosanz, T. Fayard (2003). *The CHAMP-only Earth Gravity Field Model EIGEN-2*, Advances in Space Research, Vol. 31, No. 8, p. 1883 - 1888

Reigber, C., H. Jochmann, J. Wünsch, S. Petrovic, P. Schwintzer, F. Barthelmes, K.-H. Neumayer, R. König, C. Förste, G. Balmino, R. Biancale, J.-M. Lemoine, S. Loyer, F. Perosanz (2005a). *Earth gravity field and seasonal variability from CHAMP*. In: Reigber, C., H. Lühr, P. Schwintzer, J. Wickert (eds.), *Earth Observation with CHAMP, Results from Three Years in Orbit*, p. 25-30, Springer

Reigber, C., R. Schmidt, F. Flechtner, R. König, U. Meyer, K.-H. Neumayer, P. Schwintzer, S.Y. Zhu (2005b). *An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S*, Journal of Geodynamics, 39, p. 1-10

Reubelt, T., M. Götzelmann, E.W. Grafarend (2004). *A new CHAMP gravitational field model based on the GIS acceleration approach and two years of kinematic CHAMP data*, Proceedings of the Joint CHAMP/GRACE Science Meeting, July 2004, Potsdam

Rummel, R. and O.L. Colombo (1985). *Gravity field determination from satellite gradiometry*, Bulletin Géodésique, 59, p. 233-246

Rummel, R., G. Balmino, J. Johannessen, P. Visser, P. Woodworth (2002). *Dedicated gravity field missions - principle and aims*, Journal of Geodynamics, 33, p. 3-20

Sahin, M., P.A. Cross, P.C. Sellers (1992). *Variance component estimation applied to satellite laser ranging*, Bulletin Géodésique, 66, p. 284-295

Schaffrin, B. (1985). *Das Geodätische datum mit stochastischer vorinformation*, habilitationsschrift, Deutsche Geodätische Kommission, C-313, Munich

Schaffrin, B. (1986). *On Robust Collocation*, Proc. of the First Hotine-Marussi Symp. on Math. Geodesy, p. 343-361

Schaffrin, B. (1987). *Less sensitive tests by introducing stochastic linear hypotheses*, Proc. Second International Tampere Conference in Statistics, Univ. of Tampere, p. 647-664

Schaffrin, B. (1995). *A generalized Lagrange function approach to include fiducial constraints*, Zeitschrift für Vermessungswesen 120:7, p. 325 - 333

Schaffrin, B. (1999). *Softly unbiased estimation, part 1: The Gauss-Markov model*, Linear Algebra and its applications, 289, p. 285-296

Schaffrin, B. and H.B. Iz (2001). *Integrating heterogeneous data sets with partial inconsistencies*. In: Siderius, M. (Ed.), *Gravity, Geoid and Geodynamics 2000*, Springer, p. 49-54

Schaffrin, B. (2007). *Minimum mean squared error (MSE) adjustment and the optimal Tykhonov-Philips regularization parameter via reproducing best invariant quadratic uniformly unbiased estimates (repro-BIQUUE)*, Journal of Geodesy, 82, p. 113-121

Schlatter, A. (2006). *Das neue Landeshöhennetz der Schweiz LHN95*, Dissertation Nr. 16840, ETH Zürich, IGP, in German

Schlatter, A. and U. Marti (2002). *Neues Landeshöhennetz der Schweiz LHN95*, Vermessung, Photogrammetrie, Kulturtechnik, 1/2002, p. 13-17

Schrama, E.J.O., B. Wouters, D.A. Lavellée (2007). *Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed mass variations*, Journal of Geophysical Research, vol. 112, B08407, doi: 10.1029/2006JB004882

Schwarz, K.P., M.G. Sideris, R. Forsberg (1987). *Orthometric heights without leveling*, Journal of Surveying Engineering, vol. 113, no. 1, p. 28-40

Schwintzer, P. (1990). *Sensitivity analysis in least squares gravity modelling by means of redundancy decomposition of stochastic prior information*, Internal Report, Deutsches Geodätisches Forschungsinstitut

Schwintzer, P., C. Reigber, A. Bode, Z. Kang, S.Y. Zhu, F.-H. Massmann, J.C. Raimondo, R. Biancale, G. Balmino, J.M. Lemoine, B. Moynet, J.C. Marty, F. Barlier, Y. Boudon (1997). *Long-wavelength global gravity field models: GRIM4-S4, GRIM4-C4*, Journal of Geodesy, 71, p. 189-208

Searle, S.R., G. Casella, C.E. McCulloch (1992). *Variance components*, John Wiley, New York

Seo, K.W., C.R. Wilson, S.C. Han, D.E. Waliser (2008). *Gravity Recovery and Climate Experiment (GRACE) alias error from ocean tides*, Journal of Geophysical Research, vol. 113, B03405, doi: 10.1029/2006JB004747

Sharifi, M.A., N. Sneeuw, W. Keller (2007). *Gravity recovery capability of four generic satellite formations*. In: *Harita Dergisi, proceedings of the 1st International Symposium of the International Gravity Field Service, Gravity Field of the Earth*, 28 August - 1 September 2006, Istanbul, p. 211-216

Shepperd, S.W. (1982). *A recursive algorithm for evaluating Molodenskii-type truncation error coefficients at altitude*, Bulletin Géodésique, 56, p. 95-105

Sjöberg, L.E. (1983). *Unbiased estimation of variance-covariance components in condition adjustment with unknowns - A MINQUE approach*, Zeitschrift für Vermessungswesen 108:9, p.382-387

Sjöberg, L.E. (1984). *Non-negative variance component estimation in the Gauss-Helmert adjustment model*, Manuscripta Geodaetica 9, p. 247-280

Smith, D.A. (1998). *There is no such thing as "The" EGM96 geoid: Subtle points on the use of a global geopotential model*, IGeS Bulletin No. 8, International Geoid Service, Milan, p. 17-28

Smith, W.H.F. and P. Wessel (1990). *Gridding with continuous curvature splines in tension*, Geophysics, 55, 3, p. 293-305

Sneeuw, N. and H. Schaub (2005). *Satellite clusters for future gravity field missions*. In: Jekeli, C., L. Bastos, J. Fernandes (eds.), *Gravity, Geoid and Space Missions*, IAG Symposia, vol. 129, Springer, p. 12-17

Sneeuw, N., J. Flury, R. Rummel (2005). *Science requirements on future missions and simulated mission scenarios*, Earth, Moon, and Planets, 94, p. 113-142, doi:10.1007/s11038-004-7605-x

Stahel, W.A. and A. Welsh (1997). *Approaches to robust estimation in the simplest variance components model*, Journal of Statistical Planning and Inference, 57, p. 295-319

Strang, G. (1988). *Linear algebra and its applications*, third edition, Harcourt Brace Jovanovich, Orlando

Švehla, D. and M. Rothacher (2003). *Kinematic and reduced-dynamic precise orbit determination of low earth orbiters*, Advances in Geosciences 1, p. 47-56

Swenson, S. and J. Wahr (2002). *Estimated effects of the vertical structure of atmospheric mass on the time-variable geoid*, Journal of Geophysical Research, 107, B9, 2194, doi: 10.1029/2000JB000024

Tapley, B.D., M.M. Watkins, J.C. Ries, G.W. Davis, R.J. Eanes, S.R. Poole, H.J. Rim, B.E. Schutz, C.K. Shum, R.S. Nerem, F.J. Lerch, J.A. Marshall, S.M. Klosko, N.K. Pavlis, R.G. Williamson (1996). *The Joint Gravity Model 3*, Journal of Geophysical Research, 101, B12, p. 28029 - 28049

Tapley, B.D., C.K. Shum, J.C. Ries, S.R. Poole, P.A.M. Abusali, S.V. Bettadpur, R.J. Earnes, M.C. Kim, H.J. Rim, B.E. Schutz (1997). *The TEG-3 Geopotential model*. In: Segawa, J., H. Fujimoto, S. Okubo (eds.), *Gravity, Geoid and Marine Geodesy*, IAG Symposia, vol. 117, Springer, p. 453-460

Tapley, B.D., S. Bettadpur, J.C. Ries, P.F. Thompson, M.M. Watkins (2004a). *GRACE measurements of Mass Variability in the Earth System*, Science, 305, p. 503-505

Tapley, B.D., S. Bettadpur, M. Watkins, C. Reigber (2004b). *The gravity recovery and climate experiment: Mission overview and early results*, Geophysical Research Letters, 31, L09607, doi:10.1029/2004GL019920

Tapley, B.D., J. Ries, S. Bettadpur, D. Chambers, M. Cheng, F. Condi, B. Gunter, Z. Kang, P. Nagel, R. Pastor, T. Pekker, S. Poole, F. Wang (2005). *GGM02 - An improved Earth gravity field model from GRACE*, Journal of Geodesy, 79, p. 467-478

Teunissen, P.J.G. (1988). *Towards a least-squares framework for adjusting and testing of both functional and stochastic model*, Geodetic Computing Centre, Delft, MGP Series No. 26, 2004 (reprint 1988)

Teunissen, P.J.G. (2000a). *Adjustment Theory, an introduction*, first edition, Series on Mathematical Geodesy and Positioning, Delft University Press, Delft

Teunissen, P.J.G. (2000b). *Testing Theory, an introduction*, first edition, Series on Mathematical Geodesy and Positions, Delft University Press, Delft

Teunissen, P.J.G. and A.R. Amiri-Simkooei (2008). *Least-squares variance component estimation*, Journal of Geodesy, 80, p. 65-82

Teunissen, P.J.G., D.G. Simons, C.C.J.M. Tiberius (2005). *Probability and observation theory*, Faculty of Aerospace Engineering, Delft University of Technology, lecture notes AE2-E01

Tikhonov, A.N. and V.Y. Arsenin (1977). *Solutions of ill-posed problems*, Scripta Series in Mathematics, Winston and Sons

Tscherning, C.C. and R.H. Rapp (1974). *Closed Covariance Expressions for Gravity Anomalies, Geoid Undulations, and Deflections of the Vertical Implied by Anomaly Degree Variances*, OSU report no. 208, Ohio State University, Columbus

Van Loon, J.P. (2007). *Robust estimation and robust re-weighting in satellite gravity modelling*. In: Xu. P, J. Liu, A. Dermanis (eds.). *VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy*, IAG Symposia, vol. 132, Springer, p. 43-48

Van Loon, J.P. and J. Kusche (2005). *Stochastic model validation of satellite gravity data: A test with CHAMP pseudo-observations*. In: Jekeli, C., L. Bastos, J. Fernandes (eds.), *Gravity, Geoid and Space Missions*, IAG Symposia, vol. 129, Springer, p. 24-29

Van Loon, J.P. and J. Kusche (2007). *Towards an optimal combination of satellite data and prior information*. In: Tregoning, P. and C. Rizos (eds.). *Dynamic Planet, Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools*, IAG Symposia, vol. 130, Springer, p. 345-353

Velicogna, I. and J. Wahr (2006). *Measurements of time-variable gravity show mass loss in Antartica*, Science, vo. 31, p. 1754-1756

Velicogna, I., J. Wahr, H. van den Dool (2001). *Can surface pressure be used to remove atmospheric contributions from GRACE with sufficient accuracy to recover hydrological signals?*, Journal of Geophysical Research, 106, B8, p. 16,415-16,434

Vossepoel, F.C. (2007). *Uncertainties in the mean ocean dynamic topography before the launch of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE).*, Journal of Geophysical Research, 112, C05010

Wahr, J., M. Molenaar, F. Bryan (1998). *Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE*, Journal of Geophysical Research, 103, B12, p. 30205-30229

Wu, X., M.B. Heflin, E.R. Ivins, I. Fukumori (2006). *Seasonal and interannual global surface mass variations from multisatellite geodetic data*, Journal of Geophysical Research, 111, B09401, doi:10.1029/2005JB004100

Xu, P. (1989). *On robust estimation with correlated observation*, Bulletin Géodésique, 63, p. 237-252

Xu, P. (1992). *The value of minimum norm estimation of geopotential fields*, Geophysical Journal International, 111, p. 170-178

Xu, P., Y. Shen, Y. Fukuda, Y. Liu (2006). *Variance component estimation in linear inverse ill-posed models*, Journal of Geodesy, 80, p. 69-81

Yang, Y. (1994). *Robust estimation for dependent observations*, Manuscripta Geodaetica, 19, p. 10-17

Yang, Y., M.K. Cheng, C.K. Shum, B.D. Tapley (1999). *Robust estimation of systematic errors of satellite laser range*, Journal of Geodesy, 73, p. 345-349

Yang, Y., L.J. Song, T.H. Xu (2002). *Robust estimator for correlated observations based on bifactor equivalent weights*, Journal of Geodesy, 76, p. 353-358

Yang, Y., T.H. Xu, L.J. Song (2005). *Robust estimation of variance components with application in Global Positioning System network adjustment*, Journal of Surveying Engineering, November 2005, p. 107-112

Yuan, D.N. (1991). *The determination and error assessment of the Earth's gravity field*, CSR-91-1, dissertation, Center for Space Research, The University of Texas, Austin

A

Series expansion into spherical harmonics

A harmonic function V is a scalar function, which fulfills Laplace equation:

$$\Delta V = 0, \tag{A.1}$$

in which the *Laplace* operator Δ, expressed in Cartesian coordinates, is defined as

$$\Delta V := \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \tag{A.2}$$

As

$$\Delta \left(\frac{1}{l} \right) = 0, \tag{A.3}$$

we can show that the gravitational potential is a harmonic function outside its masses, i.e.,

$$\Delta V(P) = G \int_{\Sigma_Q} \rho(Q) \Delta \left(\frac{1}{l_{PQ}} \right) \mathrm{d}\Sigma_Q = 0. \tag{A.4}$$

In Earth sciences, it is much more convenient to work with spherical (or ellipsoidal) coordinates instead of Cartesian coordinates. Transformation between the spherical coordinates (ϑ, λ, r) and the Cartesian coordinates (x, y, z) is as follows:

$$\begin{aligned} x &= r \sin\vartheta \cos\lambda, \\ y &= r \sin\vartheta \sin\lambda, \\ z &= r \cos\vartheta, \end{aligned} \tag{A.5}$$

and inversely

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2}, \\ \vartheta &= \arctan \frac{\sqrt{x^2 + y^2}}{z}, \\ \lambda &= \arctan \frac{y}{x}. \end{aligned} \tag{A.6}$$

In Heiskanen and Moritz [1967] it was shown that ΔV can be written in spherical coordinates as

$$r^2\frac{\partial^2 V}{\partial r^2} + 2r\frac{\partial V}{\partial r} + \cot\vartheta\frac{\partial V}{\partial \vartheta} + \frac{1}{\sin^2\vartheta}\frac{\partial^2 V}{\partial \lambda^2} = 0. \tag{A.7}$$

The solution to this equation reads

$$V(\vartheta,\lambda,r) = \sum_{l=0}^{\infty}\frac{1}{r^{(l+1)}}\sum_{m=0}^{l}\left[a_{lm}P_{lm}(\cos\vartheta)\cos m\lambda + b_{lm}P_{lm}(\cos\vartheta)\sin m\lambda\right]. \tag{A.8}$$

In global gravity field modelling, we use a slightly modified series expansion, i.e.

$$V(\vartheta,\lambda,r) = \frac{GM}{R}\sum_{l=0}^{\infty}\left(\frac{R}{r}\right)^{(l+1)}\sum_{m=0}^{l}\left[\bar{C}_{lm}\cos m\lambda + \bar{S}_{lm}\sin m\lambda\right]\bar{P}_{lm}(\cos\vartheta), \tag{A.9}$$

with GM the gravity-mass constant of the geopotential model, R a scale factor, $\bar{C}_{lm}, \bar{S}_{lm}$ the 4π-normalized potential coefficients and $\bar{P}_{lm}(\cos\vartheta)$ the 4π-normalized Legendre functions. The unnormalized Legendre functions can be obtained by Heiskanen and Moritz [1967]

$$P_{lm}(t) = \frac{1}{2^l l!}(1-t^2)^{m/2}\frac{\mathrm{d}^{l+m}}{\mathrm{d}t^{l+m}}(t^2-1)^l. \tag{A.10}$$

This can be split into two parts, if we first compute the Legendre polynomial

$$P_l(t) = P_{l0}(t) = \frac{1}{2^l l!}\frac{\mathrm{d}^l}{\mathrm{d}t^l}(t^2-1)^l, \tag{A.11}$$

and in a second step compute the Legendre function

$$P_{lm}(t) = (1-t^2)^{m/2}\frac{\mathrm{d}^m P_l(t)}{\mathrm{d}t^m}. \tag{A.12}$$

The Legendre polynomials are often obtained by the recursive equation

$$P_l(t) = -\frac{l-1}{l}P_{l-2}(t) + \frac{2l-1}{l}tP_{l-1}(t) \quad ; \quad P_0(t) = 1 \quad ; \quad P_1(t) = t. \tag{A.13}$$

In global gravity field determination, we make use of the 4π-normalized Legendre functions, i.e.,

$$\bar{P}_{lm}(\cos\vartheta) = H_{lm}P_{lm}(\cos\vartheta) \tag{A.14}$$

with

$$\begin{aligned} H_{lm} &= \sqrt{2l+1} && \text{for } m = 0 \\ H_{lm} &= \sqrt{2(2l+1)\tfrac{(l-m)!}{(l+m)!}} && \text{for } m \neq 0. \end{aligned} \tag{A.15}$$

The 4π-normalized Legendre functions have the orthogonality relationship

$$\int_\sigma \bar{P}_{lm}(\cos\vartheta)\bar{P}_{nk}(\cos\vartheta)\left\{\begin{array}{c}\cos m\lambda\cos k\lambda\\ \sin m\lambda\sin k\lambda\\ \cos m\lambda\sin k\lambda\\ \sin m\lambda\cos k\lambda\end{array}\right\}\mathrm{d}\sigma = \left\{\begin{array}{c}4\pi\delta_{ln}\delta_{mk}\\ 4\pi\delta_{ln}\delta_{mk}\\ 0\\ 0\end{array}\right\}. \tag{A.16}$$

B

Matrix algebra and matrix analysis

In this appendix, we will summarize several properties of matrices and matrix operators that will have been used in this thesis; see also Teunissen and Amiri-Simkooei [2008].

Partial derivatives of a matrix

$$\frac{\partial \text{tr}(AB)}{\partial A} = B^T \tag{B.1}$$

$$\frac{\partial \text{tr}(ABAC)}{\partial A} = (BAC)^T + (CAB)^T \tag{B.2}$$

$$\frac{\partial \text{tr}(A^T BAC)}{\partial A} = BAC + (CA^T B)^T \tag{B.3}$$

$$\frac{\partial \ln \det A}{\partial a_i} = \text{tr}(A^{-1}\frac{\partial A}{\partial a_i}) \tag{B.4}$$

$$\frac{\partial A^{-1}}{\partial a_i} = -A^{-1}\frac{\partial A}{\partial a_i}A^{-1} \tag{B.5}$$

Matrix inversions

$$(A + BDC)^{-1} = A^{-1} - A^{-1}B(I + DCA^{-1}B)^{-1}DCA^{-1} \tag{B.6}$$

The *Kronecker* product

Let A be a $m \times n$ matrix and B a $q \times r$ matrix. The $mq \times nr$ matrix

$$A \otimes B := \begin{bmatrix} A_{11}B & A_{12}B & \cdots & A_{1n}B \\ A_{21}B & A_{22}B & \cdots & A_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}B & A_{m2}B & \cdots & A_{mn}B \end{bmatrix} \tag{B.7}$$

is defined as the *Kronecker product* of A and B. Properties of this product are

$$\begin{array}{lcll} A \otimes (B \otimes C) & = & (A \otimes B) \otimes C & \\ A \otimes (B + C) & = & (A \otimes B) + (A \otimes C) & \\ (A + B) \otimes C & = & (A \otimes C) + (B \otimes C) & \\ (AB) \otimes (CD) & = & (A \otimes C)(B \otimes D) & \\ (A \otimes B)^T & = & A^T \otimes B^T & \\ (A \otimes B)^{-1} & = & A^{-1} \otimes B^{-1} & \text{for square non-singular } A \text{ and } B \text{ matrices} \end{array} \tag{B.8}$$

The vec-operator

The vec-operator stacks the columns of a matrix below one another. With A defined as a $m \times n$ matrix with $A = [\underline{a}_1 \underline{a}_2 \cdots \underline{a}_n]$, the $mn \times 1$ vector $\text{vec}\{A\}$ is defined as

$$\text{vec}\{A\} := \begin{bmatrix} \underline{a}_1 \\ \underline{a}_2 \\ \vdots \\ \underline{a}_n \end{bmatrix} \tag{B.9}$$

Properties of this operator are

$$\text{vec}\{ABC\} = (C^T \otimes A)\text{vec}\{B\} \tag{B.10}$$

The vh-operator

The vh-operator is obtained in quite a similar way as the vec-operator, but it starts with the diagonal elements of the square matrix A. If A is a $n \times n$ matrix, the elements of the vector $\text{vh}\{A\}$ will consist of the $n(n+1)/2$ lower triangle elements of this matrix.

The trace-operator

The tr-operator performs a summation of the diagonal elements of a square matrix. Properties of this operator are

$$\begin{array}{lcl} \text{tr}(AB) & = & \text{tr}(BA) \\ \text{tr}(A \otimes B) & = & \text{tr}(A)\text{tr}(B) \\ \text{tr}(ABCD) & = & (\text{vec}\{D^T\})^T (C^T \otimes A)\text{vec}\{B\} \end{array} \tag{B.11}$$

C

Some standard distributions

Normal distribution

The *central limit theorem* states that the summation of p different elementary errors, each with an arbitrary distribution, converges to a normal distribution if p approximates infinity. As measurement errors are, in general, a summation of multiple independent error sources, most measurement errors can be assumed to be normally distributed. The univariate normal distribution can be written as

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \tag{C.1}$$

also denoted as

$$x \sim N(\mu, \sigma). \tag{C.2}$$

This can be generalized to the multivariate normal distribution of the $n \times 1$ vector $\underline{x}$:

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2}(\text{det}Q_x)^{1/2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T Q_x^{-1}(\underline{x}-\underline{\mu})}, \tag{C.3}$$

denoted as

$$\underline{x} \sim N(\underline{\mu}, Q_x), \tag{C.4}$$

where

$$E\{\underline{x}\} = \underline{\mu} \quad ; \quad D\{\underline{x}\} = Q_x.$$

Let $\underline{y} = A\underline{x} + \underline{c}$, then

$$\underline{y} \sim N(A\underline{\mu} + \underline{c}, AQ_xA^T) \tag{C.5}$$

The relationship between the critical values and the two-sided levels of significance is shown in table C.1.

Tab. C.1: Critical values κ_α for some two-sided levels of significance α within the normal distribution

α	κ_α
0.250	1.150
0.100	1.645
0.050	1.960
0.025	2.241
0.010	2.576
0.005	2.807
0.001	3.291

Non-central χ^2-distribution

If $\underline{x} \sim N(\underline{\mu}, Q_x)$ and $y = \underline{x}^T Q_x^{-1} \underline{x}$ then $y \sim \chi^2(n, \lambda)$, with n, the degrees of freedom and λ the non-centrality parameter, defined as $\lambda = \underline{\mu}^T Q_x^{-1} \underline{\mu}$. The expectation and dispersion of y are

$$E\{y\} = n + \lambda \quad ; \quad D\{y\} = 2n + 4\lambda \tag{C.6}$$

If $\lambda = 0$ the distribution is called the central χ^2-distribution and noted as $y \sim \chi^2(n)$.

Non-central F-distribution

If $\underline{x} \sim N(\underline{\mu}, Q_x)$ and $\underline{y} \sim N(\underline{0}, Q_y)$ and $\underline{x}$ and $\underline{y}$ are uncorrelated, then

$$z = \frac{(\underline{x}^T Q_x^{-1} \underline{x})/n}{(\underline{y}^T Q_y^{-1} \underline{y})/m} \sim F(n, m, \lambda) \tag{C.7}$$

with n the dimension of $\underline{x}$, m the dimension of $\underline{y}$ and λ the non-centrality parameter defined as $\underline{\mu}^T Q_x^{-1} \underline{\mu}$. Note that the distribution of $u = (\underline{x}^T Q_x^{-1} \underline{x})/n$ is sometimes noted as $u \sim F(n, \infty, \lambda)$. If $\lambda = 0$, the distribution is called the central F-distribution and noted as $z \sim F(n, m)$.

Student's t-distribution

If $\underline{x} \sim N(\underline{0}, 1)$ and $y \sim \chi^2(k)$, then

$$\underline{z} = \frac{\underline{x}}{\sqrt{y/k}} \sim t(k) \tag{C.8}$$

with k the degrees of freedom and $t(.)$ the notation of the Student's t-distribution. Note that

$$u = \underline{z}^T \underline{z} \sim F(1, k) \tag{C.9}$$

Summary

Functional and stochastic modelling of satellite gravity data

The gravitational field of the Earth is caused by the heterogeneous mass distribution of the Earth. Due to this field, the mean sea level (at rest) shows fluctuations (geoid heights) up to 100 meters with respect to a reference ellipsoid. The time-varying gravity signal is a source of information on the melting of the polar ice, the hydrological cycle and other mass variations within the Earth. Moreover, the gravity field of the Earth is needed to derive the physical (orthometric) heights and to estimate the ocean at rest, which is necessary for oceanographic research.

The gravitational field is an harmonic field and can be expressed in a model of spherical harmonic coefficients. A higher resolution can be obtained by an increase in this number of coefficients; e.g., more than 130,000 coefficients are needed for a resolution of 111 kilometers. Traditional techniques for the determination of this gravity field are terrestrial measurements, marine measurements and airborne gravimetry. Since the launch of the first satellite in 1957, satellite tracking measurements are used to derive the gravitational field, together with altimetry measurements to estimate the sea surface at rest. With the launch of the dedicated satellite missions CHAMP (2000), GRACE (2002) and the upcoming GOCE mission (2008), millions of observations have become available, which will (and allready have) improved the gravitational field by several orders. Moreover, the measurements of the GRACE mission enable us to derive monthly estimates of the mass changes in a thin shell near the Earth's surface, mainly due to hydrology, post-glacial rebound and the melting of the polar ice. The combination of different observation groups (satellite / surface gravity data) raises a number of research questions:

How to deal with systematic effects in the data?

The (linear) relationship between the observations and the unknown parameters is defined in the functional model. These parameters consist of the 'global' coefficients of the gravitational field and some 'local' parameters, such as systematic effects. The residuals, i.e., observations minus model, should only consist of observation noise. Local parameters need to be tested for significance. A correct stochastic model is needed for

such an hypothesis testing. As unmodelled systematic effects enter into the stochastic model, which will change this stochastic model with each improvement of the vector of unknowns, this testing should be done in an iterative process.

How to weight the different observation groups?

The quality of the satellite measurements vary considerably in time due to changing GPS constellations and instrument conditions. The weighted least-squares approach will give more weight to high-quality observations. The precision of the observations is expressed in the variance-covariance matrix, as part of the stochastic model. A-priori estimates of these stochastic properties are often not reliable and in general too optimistic. To achieve precise and reliable estimates of the unknown parameters, it is therefore necessary to calibrate the stochastic model. In the 1970s and 1980s such a calibration was mainly done externally, i.e., by comparing the result to independent data. A breakthrough was the implementation of the subset solution method, which is an internal calibration method suggested by F.J. Lerch [1989]. However, with the launch of the CHAMP and GRACE missions, which have brought millions of gravity-related observations to the geodetic community, this community has made a step backwards with respect to the validation of the stochastic model. This is probably due to the huge dimensions of the least-squares problem. In this thesis, we have shown that even for large amounts of data (and unknown parameters), it is still possible to compute unbiased variance components and consequently observation weights . With the use of Monte Carlo simulations, we have re-written the equations in such a way that every least-squares software package can be used to compute the new weights of the different data sets. In this way, the stochastic model can absorb the varying quality of the satellite data and it enables an improved combination of different observation types.

How to decrease the influence of the blunders in the data?

Outliers are observations that do not follow the assumed functional and stochastic model. The residual of an outlier is much higher than one could expect from the stochastic model. The conventional approach is to test a bias in an observation for significance (w-test statistic) and consequently remove the observation with the highest test statistic (data snooping). An alternative method is to assume a different distribution than the normal distribution as the probability density function of the observations (M-estimation). In this way, more probability is assumed for observations with a high test statistic. Therefore, the influence of these observations is highly reduced. With a sufficiently high redundancy in the observations, it is even possible to estimate this probability density function from the observations and consequently re-weight all observations accordingly. This cost function estimation(CFE) produced the best results in a test setup with real CHAMP satellite gravity data. It is recommended to further refine this method and to use it on different sources of data, including the very precise measurements of the GOCE satellite.

The testing of the local parameters, the validation of the stochastic model and the detection and treatment of the outliers are all connected to each other. It is therefore recommended to perform these algorithms in an iterative process. The algorithms have been tested on several different applications in this thesis. In the first application a global gravity field model is derived from CHAMP pseudo-observations, using the energy balance approach. In the second test setup, we have computed a joint inversion of weekly GPS site displacements with monthly GRACE models to obtain estimates of the surface mass redistributions. Simulated GRACE observations are used in the third application to quantify and reduce the effect of temporal aliasing. In the last application a corrector surface to the gravimetric geoid is estimated for Switzerland using a combination of GPS/levelling data and the gravimetric geoid.

Jasper van Loon

Samenvatting

Functional and stochastic modelling of satellite gravity data

Het zwaartekrachtsveld van de Aarde wordt veroorzaakt door de massaverdeling binnen de Aarde en heeft tot gevolg dat het zeeoppervlak in rust afwijkingen (geoïdehoogten) vertoont tot wel 100 meter ten opzichte van een referentie-ellipsoïde. Veranderingen in dit zwaartekrachtsveld geven ons zo informatie over de smelting van de poolkappen, de hydrologische cyclus en de massaveranderingen in het binnenste van de Aarde. Bovendien is het zwaartekrachtsveld essentieel in de bepaling van de fysische (orthometrische) hoogte en de definiëring van het zeeoppervlak in rust, hetgeen nodig is voor oceanografisch onderzoek.

Dit zwaartekrachtsveld is een harmonisch veld en kan uitgedrukt worden in een model van spherisch harmonische coefficiënten. Een verhoging van dit aantal coefficiënten leidt tot een verbetering van de resolutie. Zo heeft men voor een resolutie van 111 kilometer meer dan 130.000 coefficiënten nodig. Traditionele technieken voor de bepaling van dit zwaartekrachtsveld zijn zwaartekrachtmetingen op land, mariene metingen en vliegtuiggravimetrie. Sinds het begin van het ruimtetijdperk (1957) worden ook afstandsmetingen naar satellieten uitgevoerd, alsmede altimetriemetingen van het zeeoppervlak. Met de lancering van de satellietmissies CHAMP (2000), GRACE (2002) en de toekomstige GOCE missie (2008) komen vele miljoenen metingen beschikbaar, die ervoor zorgen (en al hebben gezorgd) dat het zwaartekrachtsveld met enkele ordes beter bepaald kan worden. De waarnemingen van de GRACE-missie maken het bovendien mogelijk om voor het eerst op een maandelijkse basis de massaveranderingen in de bovenste laag van de Aarde te berekenen. Deze massaverplaatsingen zijn voornamelijk toe te schrijven aan hydrologie, post-glaciale opheffing en het smelten van de poolkappen. De combinatie van verschillende waarnemingsgroepen (satellietmetingen / terrestrische metingen) levert een aantal onderzoeksvragen op:

Hoe om te gaan met de systematische effecten per data set?

Het functionaal model geeft de relatie weer tussen de waarnemingen en de onbekende parameters. Deze onbekende parameters bestaan uit de 'globale' coefficiënten van het zwaartekrachtsveld en enkele 'locale' parameters, zoals systematische fouten.

De residuen, oftewel waarnemingen minus model, dienen idealiter slechts de ruis te bevatten. De locale coefficiënten dienen getest te worden op significantie. Hiervoor is echter ook het stochastische model nodig. Aangezien ongemodelleerde systematische fouten geabsorbeerd worden door het stochastic model, zal dit model bij iedere schatting van de onbekende parameters aangepast moeten worden. Hierdoor is het testen van de locale parameters een iteratief proces.

Hoe moeten de verschillende waarnemingsgroepen gewogen worden?

De kwaliteit van satellietmetingen verandert sterk in de tijd door veranderende GPS-constellaties en het functioneren van de meetapparatuur. In een gewogen kleinste-kwadraten oplossing krijgen metingen met een hogere kwaliteit een hoger gewicht in de schatting van de onbekende parameters. De kwaliteit wordt uitgedrukt in de variantie-covariantie matrix, als onderdeel van het stochastisch model van de waarnemingen. De a priori schattingen van deze stochastische eigenschappen zijn meestal niet betrouwbaar en over het algemeen te positief. Een calibratie van het stochastisch model is daardoor noodzakelijk voor een precieze en betrouwbare schatting van de onbekende parameters. In de jaren '70 en '80 werd deze calibratie voornamelijk uitgevoerd door het eindproduct te vergelijken met onafhankelijke gegevens. Een doorbraak was de 'subset solution'-methode, een interne calibratiemethode, voorgesteld door F.J. Lerch [198i9]. Met de komst van miljoenen waarnemingen van de nieuwe satellietmissies CHAMP en GRACE heeft de geodetische wereld echter een stap teruggedaan voor wat betreft de validatie van het stochastische model. Dit is waarschijnlijk toe te schrijven aan de enorme dimensiteit van het kleinste-kwadraten probleem. In dit proefschrift laten we echter zien dat ook voor grote hoeveelheden data het mogelijk is om de zuivere variantiecomponentenschatting toe te passen. Met behulp van Monte Carlo simulaties zijn de vergelijkingen zo opgeschreven dat ieder kleinste-kwadraten softwarepakket gebruikt kan worden voor de berekening van de nieuwe gewichten van de verschillende data sets. Op deze manier kan de variërende kwaliteit van de satellietgegevens worden opgevangen door het stochastisch model en is een betere combinatie mogelijk tussen de verschillende soorten waarnemingen.

Hoe kan de invloed van blunders zo veel mogelijk beperkt worden?

Blunders zijn waarnemingen, die niet overeenkomen met de aangenomen functionale en stochastische modellen. De residuen zijn veel groter dan verwacht zou worden, uitgaande van het stochastisch model. De conventionele aanpak is om een afwijking in de waarneming te testen voor significantie (w-toets) en vervolgens de waarneming met de hoogste testgrootheid te verwijderen uit de data set (data snooping). Een alternatieve methode is echter om een andere verdeling voor de waarnemingen aan te nemen dan de normaalverdeling. Hierdoor vermoed men meer kans op een blunder in de waarnemingen en verkleint men zo de invloed van zo'n blunder. Nog een stap verder is om deze kansverdeling rechtstreeks te schatten uit de waarnemingen. Deze 'cost function

estimation (CFE)' levert in een testopzet met echte CHAMP-waarnemingen de beste oplossing. Het verdient aanbeveling om deze methode te verfijnen en toe te passen op verschillende soorten data, waaronder de nieuwe, zeer nauwkeurige metingen van de GOCE-satelliet.

Het testen van locale parameters, het verbeteren van het stochastic model en de behandeling van blunders in de data zijn gerelateerd aan elkaar. Het is daarom aan te raden dit in een iteratief proces op te lossen. Dit is in dit proefschrift gedaan voor meerdere applicaties betreffende zwaartekrachtgegevens. Zo is het globale zwaartekrachtsveld gemodelleerd, gebruikmakend van satellietmetingen van de CHAMP-missie. De tweede applicatie vormt de combinatie van wekelijkse GPS-stationsverplaatsingen met maandelijkse GRACE-modellen voor de bepaling van massaverplaatsingen in de opperste laag van de aarde. De derde toepassing richt zich op het reduceren van het 'temporal aliasing'-effect. Hierbij is gebruik gemaakt van gesimuleerde GRACE-waarnemingen. De laatste toepassing is de bepaling van het correctieoppervlak in Zwitserland door gebruik te maken van GPS-waarnemingen, waterpasmetingen en de gravimetrische geoïde.

Jasper van Loon

Curriculum Vitae

Jasper van Loon was born in Halsteren, The Netherlands, on May 7, 1978. From 1990 to 1996, he attended secondary school at the Regionale Scholengemeenschap 't Rijks in Bergen op Zoom. After that he studied Geodetic Engineering at the Delft University of Technology. In 2000, he went on a three month internship at the Kort og Matrikelstyrelsen (KMS) in Copenhagen, where he worked on the validation of the NKG data base and on the improvement of local geoids. In 2002, he finished his MSc thesis, titled *Fast inversion of satellite gradiometry data*, supervised by Roland Klees. From September 2002 to March 2003, he worked as a Research Associate at the Delft Institute of Earth Observation and Space Systems (DEOS), directly followed by his PhD research at the same institute. The PhD research was daily supervised by Jürgen Kusche and from 2007 by Roland Klees, who was also his promotor. In 2008, Jasper started as a Postdoctoral fellow at the same institute. He will defend this thesis *Functional and stochastic modelling of satellite gravity data* on the 24th of November, 2008.

Besides his work as a researcher, Jasper has been active as a volunteer in several organisations. He has been a leader and board member of Scouting Halsteren, youth leader at the volleyball club Kratos '08 in Delft and organizer of several trips to South Africa as part of the North-South partnership of Scouting Nederland.