

EARTH GRAVITY FIELD FROM SPACE – FROM SENSORS TO EARTH SCIENCES

Cover illustration: Since the time of Newton the falling apple is regarded as a symbol of gravitation. Nowadays it symbolizes the free fall of satellites guided by the Earth's gravity field in their orbit around the Earth. The highly exaggerated undulations of the geoid surface demonstrate the variations of the Earth's gravity field. They are caused by density contrasts inside the Earth and at its surface.

Space Sciences Series of ISSI

Volume 17

The International Space Science Institute is organized as a foundation under Swiss law. It is funded through recurrent contributions from the European Space Agency, the Swiss Confederation, the Swiss National Science Foundation, and the University of Bern. For more information, see the homepage at <http://www.issi.unibe.ch/>.

The titles published in this series are listed at the end of this volume.

EARTH GRAVITY FIELD FROM SPACE – FROM SENSORS TO EARTH SCIENCES

*Proceedings of an ISSI Workshop
11–15 March 2002, Bern, Switzerland*

Edited by

G. BEUTLER

Astronomical Institute, University of Bern, Bern, Switzerland

M.R. DRINKWATER

European Space Agency, ESTEC, Noordwijk, The Netherlands

R. RUMMEL

Institute of Astronomical and Physical Geodesy, TU München, Germany

R. VON STEIGER

International Space Science Institute, Bern, Switzerland

Reprinted from *Space Science Reviews*, Volume 108, Nos. 1–2, 2003



Springer-Science+Business Media, B.V.

A.C.I.P. Catalogue record for this book is available from the Library of Congress

ISBN 978-90-481-6334-2 ISBN 978-94-017-1333-7 (eBook)
DOI 10.1007/978-94-017-1333-7

Printed on acid-free paper

All Rights Reserved

© 2003 Springer Science+Business Media Dordrecht
Originally published by Kluwer Academic Publishers in 2003

Softcover reprint of the hardcover 1st edition 2003

No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner

TABLE OF CONTENTS

Foreword	x
R. RUMMEL / How to Climb the Gravity Wall	1
I: PRECISE ORBIT DETERMINATION AND GRAVITY FIELD MODELLING	
U. HUGENTOBLER and G. BEUTLER / Strategies for Precise Orbit Determination of Low Earth Orbiters Using the GPS	17
P.N.A.M. VISSER and J. VANDEN IJSSEL / Aiming at a 1-cm Orbit for Low Earth Orbiters: Reduced-Dynamic and Kinematic Precise Orbit Determination	27
N. SNEEUW / Space-Wise, Time-Wise, Torus and Rosborough Rep- resentations in Gravity Field Modelling	37
G. BALMINO / Gravity Field Recovery From GRACE: Unique Aspects of the High Precision Inter-Satellite Data and Analysis Methods	47
C. REIGBER, G. BALMINO, P. SCHWINTZER, R. BIANCALE, A. BODE, J.-M. LEMOINE, R. KÖNIG, S. LOYER, H. NEUMAY- ER, J.-C. MARTY, F. BARTHELMES, F. PEROSANZ and S.Y. ZHU / Global Gravity Field Recovery Using Solely GPS Track- ing and Accelerometer Data from CHAMP	55
W.-D. SCHUH / The Processing of Band-Limited Measurements: Filter- ing Techniques in the Least Squares Context and in the Presence of Data Gaps	67
II: SOLID EARTH PHYSICS	
M.E. TAMISIEA, J.X. MITROVICA, J.L. DAVIS and G.A. MILNE / Long Wavelength Sea Level and Solid Surface Perturbations Driven by Polar Ice Mass Variations: Fingerprinting Greenland and Antarctic Ice Sheet Flux	81
A.M. MAROTTA / Benefits From GOCE within Solid Earth Geophysics	95
L.L.A. VERMEERSEN / The Potential of GOCE in Constraining the Structure of the Crust and Lithosphere from Post-Glacial Rebound	105

R. TONDI, R. DE FRANCO and G. BIELLA / Deep and Shallow Solid-Earth Structures Reconstructed with Sequential Integrated Inversion (SII) of Seismic and Gravity Data	115
A. CAZENAVE, C. CABANES, K. DOMINH, M.C. GENNERO and C. LE PROVOST / Present-Day Sea Level Change: Observations and Causes	131
III: OCEAN CIRCULATION	
C. WUNSCH and D. STAMMER / Global Ocean Data Assimilation and Geoid Measurements	147
C. LE PROVOST and M. BREMOND / Resolution Needed for an Adequate Determination of the Mean Ocean Circulation from Altimetry and an Improved Geoid	163
E.J.O. SCHRAMA / Error Characteristics Estimated from CHAMP, GRACE and GOCE Derived Geoids and from Satellite Altimetry Derived Mean Dynamic Topography	179
S. IMAWAKI, H. UCHIDA, K. ICHIKAWA and D. AMBE / Estimating the High-Resolution Mean Sea-Surface Velocity Field by Combined Use of Altimeter and Drifter Data for Geoid Model Improvement	195
K. HAINES, R. HIPKIN, C. BEGGAN, R. BINGLEY, F. HERNANDEZ, J. HOLT, T. BAKER and R.J. BINGHAM / Combined Use of Altimetry and In Situ Gravity Data for Coastal Dynamics Studies	205
C.W. HUGHES and V. STEPANOV / Feasibility, and Contribution to Ocean Circulation Studies, of Ocean Bottom Pressure Determination	217
P. LE GRAND / Impact of Geoid Improvement on Ocean Mass and Heat Transport Estimates	225
P.Y. LE TRAON, F. HERNANDEZ, M.H. RIO and F. DAVIDSON / How Operational Oceanography can Benefit from Dynamic Topography Estimates as Derived From Altimetry and Improved Geoid	239

IV: GEODESY

F. SACERDOTE and F. SANSÒ / Remarks on the Role of Height Datum in Altimetry-Gravimetry Boundary-Value Problems	253
P. KNUDSEN / Ocean Tides In GRACE Monthly Averaged Gravity Fields	261
R.D. RAY, D.D. ROWLANDS and G.D. EGBERT / Tidal Models in a New Era of Satellite Gravimetry	271
M. VERMEER / The Elusive Stationary Geoid	283
J. BOUMAN and R. KOOP / Geodetic Methods for Calibration of GRACE and GOCE	293

V: SEA LEVEL

P. WOODWORTH and J.M. GREGORY / Benefits of GRACE and GOCE to Sea Level Studies	307
J. WAHR and I. VELICOGNA / What Might GRACE Contribute to Studies of Post Glacial Reboud?	319
R.S. NEREM, J. WAHR and E.W. LEULIETTE / Measuring the Dis- tribution of Ocean Mass Using GRACE	331
S. SWENSON and J. WAHR / Monitoring Changes in Continental Water Storage with GRACE	345

VI: FUTURE CONCEPTS

E. CANUTO, P. MARTELLA and G. SECHI / Attitude and Drag Con- trol: An Application to the GOCE Satellite	357
S. ZAREMBIŃSKI / On Superconductive Gravity Gradiometry in Space	367
P.L. BENDER, J.L. HALL, J. YE and W.M. KLIPSTEIN / Satellite- Satellite Laser Links for Future Gravity Missions	377
P.L. BENDER, R.S. NEREM and J.M. WAHR / Possible Future Use of Laser Gravity Gradiometers	385
P. TOUBOUL / MICROSCOPE Instrument Development, Lessons for GOCE	393

M. AGUIRRE-MARTINEZ and N. SNEEUW / Needs and Tools for Future Gravity Measuring Missions	409
---	-----

VII: CLOSING SESSION

M.R. DRINKWATER, R. FLOBERGHAGEN, R. HAAGMANS, D. MUZI and A. POPESCU / GOCE: ESA's First Earth Explorer Core Mission	419
G. BEUTLER / Earth Gravity Field From Space – From Sensors to Earth Sciences: Closing Remarks	433
Author Index	443
List of Participants	445

FOREWORD

This volume in the *Space Sciences Series of ISSI* is dedicated to the determination of the Earth's gravity field, which was the topic of a workshop organized by ISSI from 11 to 15 March 2002 in Bern. The aim of the meeting was to gather active researchers from various fields (satellite geodesy, gravity field modelling, solid Earth physics, oceanography, etc.) to assess the exciting new developments that are made possible by space missions currently in orbit or about to be launched.

Following some general enquiries with the scientific community concerning the desirability of organising a workshop on the proposed topic, a group of four convenors (Gerhard Beutler, Univ. of Bern; Mark Drinkwater, ESA-ESTEC, Noordwijk; Reiner Rummel, Techn. Univ. München; Rudolf von Steiger, ISSI, Bern) formulated the aims and goals of the workshop, structured the workshop into six sections, nominated a list of session chairs (Georges Balmino, GRGS-CNRS, Toulouse; Roberto Sabadini, Univ. of Milano, Christian Le Provost, Obs. Midi-Pyrénées, Toulouse; Philip Woodworth, Bidston Observatory, Birkenhead; Miguel Aguirre, ESA-ESTEC, Noordwijk), who in turn set up a programme of introductory talks. The workshop was organized such as to have only plenary sessions with typically half-hour presentations and ample time for discussions. The last day was devoted to conclusions and future objectives.

The ESA Earth science explorer core mission GOCE, to be launched in 2006, will enhance our knowledge of the global static gravity field and of the geoid by orders of magnitude. The U.S. satellite gravity mission GRACE (2002-2006) is currently measuring, in addition, the temporal variations of the gravity field. With these new data a whole range of fascinating new possibilities will be opened for solid Earth physics, oceanography, geodesy and sea level research. The new generation of gravity missions employs sensor concepts for gravity field measurement, orbit and attitude control and orbit determination that show interesting similarities with space experiments planned in the field of fundamental physics. This workshop brought together some 50 acknowledged experts in their field to discuss (1) strategies for ultra precision orbit determination and gravity field modelling with the data of the upcoming gravity field missions, (2) the use of accurate and high resolution gravity models in Earth sciences whereby, in particular, synergy is expected between the various science fields in their use of this type of new information, and (3) gravity field requirements and possible sensor and mission concepts for the time after GRACE and GOCE.

This volume is the collection of the papers that resulted from the talks given at the workshop, so we think it gives a good overview of the topics and their interrelations. All papers were peer reviewed by referees.

It is a pleasure to thank all those who have contributed to this volume and to the workshop in general. We thank the authors for writing up their contributions and the reviewers for their critical reports. We also thank the directorate and staff of ISSI for selecting this topic for a workshop and for their support in making it happen, in particular Johannes Geiss, Brigitte Fasler, Gabriela Indermühle, Vittorio Manno, Xavier Schneider, and Silvia Wenger. Last but not least we owe particular thanks to Ursula Pfander for her invaluable and skillful assistance in editing this volume.

March 2003

G. Beutler, M.R. Drinkwater, R. Rummel, R. von Steiger



Group Photograph



- | | | |
|--------------------|-------------------|--------------------|
| 1. N. Sneeuw | 14. S. Imawaki | 27. A. Cazenave |
| 2. P. Schwintzer | 15. F. Sansò | 28. J. Wahr |
| 3. R. Haagmans | 16. W. Spakman | 29. M.E. Tamisiea |
| 4. R. Ray | 17. G. Beutler | 30. S. Zarembinski |
| 5. P. Le Grand | 18. W.D. Schuh | 31. A.M. Marotta |
| 6. P. Knudsen | 19. B. Chao | 32. P. Visser |
| 7. Mrs. Le Provost | 20. K. Haines | 33. P. Woodworth |
| 8. E.J.O. Schrama | 21. G. Balmino | 34. M. Vermeer |
| 9. V. Toumazou | 22. J. Müller | 35. R. Sabadini |
| 10. Mrs. Imawaki | 23. D.B. Stammer | 36. R. Tondi |
| 11. M. Drinkwater | 24. S. Nerem | 37. R. Rummel |
| 12. R. Koop | 25. P. Bender | 38. J.O. Dickey |
| 13. C. Le Provost | 26. B. Vermeersen | 39. R. de Franco |

Picture taken by R. von Steiger at the Roman amphitheater in Avenches.

HOW TO CLIMB THE GRAVITY WALL

R. RUMMEL

*Institute of Astronomical and Physical Geodesy, TU München,
D-80290 München, Germany (rummel@bv.tum.de)*

Received: 27 November 2002; Accepted in final form: 16 January 2003

Abstract. What type of gravity satellite mission is required for the time after GRACE and GOCE? Essentially, the variables at our disposal are experiment altitude, compensation of attenuation by differential measurement and measurement precision. The latter depends on the performance of the complete sensor system and involves items such as required dynamic range, baseline length, sensor type (ambient temperature or cryogenic), number of test masses, etc.

Before any mission profile is to be studied the science issues to be addressed by GRACE and GOCE follow-on missions need to be clarified. Whether further improvement of the quasi-stationary part of the gravity fields is needed depends on the needs in solid earth physics, oceanography and geodesy and on the availability and quality of complementary data. Complementary data are also the key to the adequate use of measurements of temporal variations of gravity, apart from issues such as spatial and temporal data coverage.

1. A New Era of Satellite Gravity Field Mapping

Gravity field recovery by means of satellites enters currently into a new era. The first satellite of a new generation, capable of measuring the earth's gravity field with much more detail and precision, is CHAMP. It has been launched in July 2000, and combines gravity field determination with magnetic field measurements and atmospheric sounding. The gravity signal is thereby deduced from continuous and three-dimensional tracking of the CHAMP spacecraft relative to the satellites of the Global Positioning System GPS (satellite-to-satellite tracking in the high-low mode or SST h-l) in combination with micro-accelerometry. The satellite is slowly descending from a starting altitude of 454 km. The accelerometer, located at the centre of mass of the spacecraft, measures along three orthogonal axes the effect of all non-gravitational forces on the spacecraft. For information on CHAMP see www.gfz-potsdam.de. The second mission is GRACE. It is a dedicated gravity field satellite mission aiming at an even more refined gravity field recovery, and in particular, at the determination of temporal variations. It has been launched in March 2002 and consists of two almost identical satellites following each other in the same orbit, at a distance of around 200 km. The relative motion of the two spacecraft is measured with an ultra-precise microwave system in the K-band (satellite-to-satellite tracking in the low-low mode or SST l-l). The orbit altitude at the beginning of the mission was 500 km. In addition, as with CHAMP albeit more



Space Science Reviews **108:** 1–14, 2003.

© 2003 Kluwer Academic Publishers.

accurately, each satellite is equipped with a three-component micro-accelerometer at its centre of mass and with a receiver for high-low tracking of the GPS satellites, see www.csr.utexas.edu/grace/. The nominal mission duration of both, CHAMP and GRACE, is five years. In 2005 or 2006, as a third mission, GOCE will follow, like GRACE exclusively designed for high-resolution gravity field determination. Its core instrument will be a gravity gradiometer, based on the principle of differential accelerometry. In addition, GOCE will carry a GPS-receiver and it will be furnished with active drag compensation and angular control. The drag compensation will be used on the one hand to maintain GOCE at an extremely low orbit altitude (250 km), and on the other hand, to compensate for the effect of all non-gravitational forces acting on the spacecraft. The mission duration will be 18 months. For GOCE see Drinkwater *et al.* (2003). While GRACE is particularly tailored towards maximum measurement precision at medium wavelengths at the expense of some spatial resolution, in order to be able to recover time variable gravitational signals, GOCE is optimized for maximum spatial resolution at the expense of some precision.

The need for this kind of satellite projects has been pointed out as early as 1969. A workshop dealing with the required satellite missions for solid earth physics and oceanography, known as Williamstown Report (1969), contains already a remarkably comprehensive analysis of the use of more detailed and accurate gravity field knowledge for earth sciences. Subsequently, the scientific rationale for applications in geodesy, geophysics and oceanography has been consolidated and refined in a long list of studies, workshops and reports. However, it turned out to be very difficult to get any mission of this type incorporated into the programmes of one of the leading space agencies. In retrospect, one can say that the delay of more than thirty years for launching the first dedicated gravity mission was well used for strengthening the science case and for clarification of the mission concepts. Now with GRACE and GOCE, the two fundamental options, SST in the low-low mode on the one hand and gravity gradiometry on the other, are realized almost in parallel. With GRACE and GOCE underway — and assuming both missions to meet their mission objectives — it is legitimate to ask whether there is room and need for further improvement. Is there a scientific need for further improvement of gravity field knowledge and, if yes, how can it be realized by means of satellite techniques?

2. The Variables of Future Mission Concepts

We turn first to the second part of the question about the options for a next generation gravity field experiment. Let us assume the gravitational field of the earth to be

expressed in a spherical harmonic series. Then the average signal size per degree l of a spherical harmonic coefficient can be expressed by “Kaula’s rule of thumb”:

$$\frac{\sigma_l}{\sqrt{2l+1}} \approx \frac{10^{-5}}{l^2}; \quad (1)$$

(with σ_l the square root of the signal degree variance). Since the contribution of the main term, the spherical earth, is one and that of the earth’s flattening of the order 10^{-3} , we see that all deviations of the actual field from an ellipsoidal model are of the order of 10^{-5} or less. Furthermore, the spherical harmonic degree l corresponds to spatial scales s (s equal to the half wavelength) of a gravity signal like

$$s [\text{km}] = 20000 \text{ km}/l. \quad (2)$$

This implies, for example, that the signal contribution of features with $s = 200 \text{ km}$ is of the order of 10^{-9} . Already from this elementary observation one sees that it is principally very difficult to discern small scale features in the presence of the main terms of the field. This would be the situation when doing gravity field observations on the earth’s surface. At satellite altitude it gets even more difficult due to the attenuation of the field with increasing altitude. With R the earth’s mean radius and h the satellite altitude the signal size at altitude is:

$$\frac{\sigma_l(h)}{\sqrt{2l+1}} \approx \left(\frac{R}{R+h} \right)^{l+1} \cdot \frac{10^{-5}}{l^2}. \quad (3)$$

The attenuation factor decreases very quickly with increasing satellite altitude and even more with increasing degree (or decreasing spatial scales). The attenuation factor for the above example is about 4.5×10^{-2} at 200 km altitude, while it is 1×10^{-3} at 450 km .

In other words, it is intrinsically difficult to determine the earth’s gravity field from space and it is particularly difficult to recover small scale features. Since satellite orbits cannot be chosen arbitrarily low, however, there exists some fundamental limit in gravity field recovery, both in terms of precision and resolution, which cannot be overcome. Let us refer to this limit as a “gravity wall”. Given this situation it is worthwhile — in view of future missions — to think about strategies on how to climb this wall as high and as efficiently as possible.

In order to build a strategy of this kind essentially three variables are at our disposal that will be discussed in the following:

- choice of a low experiment altitude,
- compensation of attenuation by differential measurements,
- increasing the measurement precision.

2.1. LOW EXPERIMENT ALTITUDE

The best choice for any high resolution gravity satellite mission is simply the lowest possible altitude. In principle, higher altitudes may be considered as a natural low pass filter. Such low pass filtering may be employed in order to attain maximum cleanliness if very long-wavelength features are to be measured such as temporal changes in the earth's oblateness or relativistic parameters. Essentially this is the concept behind lunar laser ranging or LAGEOS observations. This rationale has also been chosen for the TIDES proposal by Colombo and Chao (1993). In all other cases the altitude must be chosen as low as possible. For a non-drag free mission with a planned life time of several years the lowest meaningful starting altitude will be around 450 km. The satellite will slowly descend due to atmospheric drag, but can be lifted in-between, as is done in the case of the CHAMP space-craft. Mission life-time will depend on solar activity and on the aerodynamic parameters of the spacecraft.

For a drag free mission the lowest possible altitude will be around 200 km. The spacecraft is actively maintained at this altitude. It requires a sophisticated compensation system. Due to the weight of the required fuel maximum mission life time is an issue. Also temporal variations of self-gravitation due to fuel consumption have to be considered.

2.2. COMPENSATION OF ATTENUATION BY DIFFERENTIAL MEASUREMENTS

The sensitivity with respect to short scale gravity signals relative to the total signal strength of large scale features can be amplified by means of differential measurements. This is the idea underlying the measurement concepts of GRACE and GOCE. In the case of SST I-I it is the measurement of the difference of the motion of the two satellites following each other (the latter being expressed in terms of velocity or acceleration differences). The measurement of the difference amplifies the contribution of smaller scale gravity features on the motion of each of the two satellites at the expense of the contribution of the large scale signals. In the case of gradiometry one aims directly at the measurements of the gravity gradients. In both cases the effect of differential measurement can be tuned through the choice of the baseline length of the sensor system. This aspect will be addressed now.

2.3. INCREASING THE MEASUREMENT PRECISION

Any advanced gravity field satellite mission is more than just a precise gravity measurement device in a satellite. Highest precisions depend on a very "clean and quiet" spacecraft environment and motion. Apart from the actual gravity sensor system it requires

- the orbit to be known very accurately,
- the effect of non-gravitational forces either being measured or compensated for,

- the orientation of the gravity sensor in space being well known,
- signal contributions due to satellite rotation either being eliminated, or measured, or controlled, and
- signal contributions due to vibrations, clanks, or time varying eigengravitation being avoided.

The gravity sensor unit may be a single instrument, such as a gradiometer or it may consist of a system of several satellites with well controlled and precise measurements connecting them.

Thus, a first conclusion is that for any advanced gravity mission “**the satellite or satellites together with a system of sensor and control elements form one gravity measuring device**”. Typical sensor elements are

- accelerometers either as a single instrument placed at the satellite’s centre of mass (as for CHAMP or GRACE) or combined to a gradiometer instrument (as for GOCE),
- in the case of a rotating gradiometer, a sensor device measuring torsions,
- a microwave or laser link between two or more satellites or directly between the proof masses of the accelerometers,
- a star tracker,
- satellite positioning based on SST h-l relative to a Global Navigation Satellite System (GNSS),
- angular control including a thruster system (and magnetic torquers), and
- drag control and/or active orbit maintenance including a thruster system.

These are the building blocks of any satellite gravity field mission. The lower the orbit and the higher the envisaged precision the more demanding are the requirements on all subsystems, as well as on the spacecraft structure and the material.

3. Attainable Measurement Precision – On What Does it Depend?

The performance of the gravity measuring device depends on a variety of factors. These are, in particular,

- the gravity sensor unit,
- the baseline length,
- the dynamic range, and
- the use of effects such as resonances. See also Table I.

3.1. GRAVITY SENSOR UNIT

For simplicity, let us denote all differential gravity sensors as gradiometers. A gradiometer is a device capable of measuring the second spatial derivative V_{ij} of the gravitational potential (or, equivalently, the first derivative of one of the components of the gravitational acceleration vector). Being a second derivative the gradient is associated to two spatial directions “ i ” and “ j ”, one defined by

the direction of the sensor baseline, the second one by the sensing direction: The second derivative is approximated by a finite difference over the length of the baseline b . Thus, from a certain baseline length on the finite difference approximation may not be accurate enough. The gradiometer may be capable of measuring all possible nine components V_{ij} , $i = 1, 2, 3$ and $j = 1, 2, 3$, or any number less than nine. Principally, gradiometer measurements are based upon the measurement of force differences. The difference of the gravitational force exerted by the mass distribution of the earth system and of sun, moon and planets on two or more test masses is the measured quantity. The force difference may be sensed as a relative motion of two test masses (two satellites or shielded test masses inside each one of two satellites). Alternatively, it may be translated into a torsion, similar to a torsion balance, then we denote it a **torsional gradiometer**, or it is measured linearly either as an **acceleration difference** between two accelerometers or as a spring force applied to a spring (mechanical, electrical or magnetic) connecting the two test masses (Wells, 1984).

A fourth alternative would be purely kinematical, the monitoring, *e.g.* by three orthogonally arranged CCD cameras, of the relative motion of test masses, subject to the earth's gravitational field, freely floating inside a spacecraft.

Remarks:

- a torsional gradiometer has been proposed by Forward (1974),
- GRACE can be viewed as a one-component gradiometer with the baseline length being the intersatellite distance,
- the GOCE gradiometer is based on differential accelerometry,
- the tethered satellite gradiometer proposed by Colombo *et al.* (1976) belongs to the category of differential accelerometry.
- a purely kinematic device has not yet been investigated. It has been suggested in Rummel (1986). From a certain perspective SST h-1 between the GPS satellites and a low orbiter belongs to this category, too.

So far, all gravity satellite missions use ambient temperature accelerometers. CHAMP, GRACE and GOCE employ different versions of the so-called STAR accelerometer by ONERA, see Touboul (2003). An alternative development is that by MACEK. A significant improvement of measurement precision would be feasible with a super-conducting gradiometer. The development of a superconducting gradiometer has been pursued for many years by Paik and Richard (1986) and co-workers and more recently by Lumley *et al.* (2002). The effect of thermal noise could be reduced significantly and internal dissipation decreased. Of course, it would require to carry a liquid helium dewar in the satellite and to cope with all related difficulties. A high temperature super-conducting gradiometer has been investigated by Zarembinski (2003). A completely new situation may arise from the development of atom interferometer-based gravity gradiometers, see *e.g.* Snadden *et al.* (1998).

3.2. BASELINE LENGTH

We take a look at a one-component gradiometer measurement Γ_{ij} with i being the direction of the baseline and j the sensing direction of the two accelerometers mounted at the end of the baseline in A and B . Then it is:

$$\Gamma_{ij} = \frac{a_j(B) - a_j(A)}{b_{AB}^i} = \frac{\Delta a_j}{b^i}. \quad (4)$$

The measurement precision is determined by the error std.dev. $\sigma(a)$ of each of the two accelerometers (assumed here to be equal) and the uncertainty of the baseline length $\sigma(b)$:

$$\frac{\sigma(\Gamma)}{\Gamma} = 2 \frac{\sigma(a)}{\Delta a} + \frac{\sigma(b)}{b} = 2 \frac{\sigma(a)}{\Gamma \cdot b} + \frac{\sigma(b)}{b}. \quad (5)$$

The size of the diagonal gravity gradient terms is $\Gamma = 3000$ E for the radial component and $\Gamma = 1500$ E for the two horizontal components. All off-diagonal components are less than 10 E, see *e.g.* Rummel and van Gelderen (1992).

For a GOCE-type gradiometer the baseline is rigid (if thermal expansion of the structure can be neglected). Then the second term in (5) is zero and the relative precision of the gradiometer is determined by the precision of the accelerometer measurement and the length of the baseline. Doubling the arm length would increase the gradiometer precision by a factor of two. The arm-length of the GOCE gradiometer is 0.5 m. The accelerometer precision is supposed to be $10^{-12} \text{ m s}^{-2}/\sqrt{\text{Hz}}$. Then we arrive at $4 \times 10^{-3} \text{ E}/\sqrt{\text{Hz}}$. The gradiometer precision inside the measurement bandwidth is $3 - 4 \times 10^{-3} \text{ E}/\sqrt{\text{Hz}}$. In the case of GRACE the baseline length of the “gradiometer” is measured by ultra-precise microwave measurements. Thus, the second term in (5) has to be maintained. However, assuming an intersatellite distance of 200 km and a precision of the distance measurement of $\sigma(b) = 1 \mu\text{m}$ this term becomes as small as 0.5×10^{-11} and can be neglected again. With $\sigma(a) = 10^{-10} \text{ m s}^{-2}\sqrt{\text{Hz}}$ (inside the measurement bandwidth) and $\Gamma = 1500$ E in along track direction one arrives at a gradiometer precision of about $10^{-6} \text{ E}/\sqrt{\text{Hz}}$. This tremendous precision is primarily the result of the large baseline length. It explains why GRACE is capable of measuring time variable gravity signals. Still, the baseline length is acceptable if one does not intend to resolve gravity features of much shorter spatial scales. Of course, these are only conceptual considerations. The actual precision depends on the performance of the complete sensor system, as already pointed out.

One may thus conclude, the increase of the baseline length is one important means to increase gradiometric performance. Naturally one has to make sure to keep the error sources due to all other sensor elements also below this level and this may be difficult to achieve.

TABLE I
Future Gravity Satellite Mission — Design Variables

orbit altitude	high: compact SAT	low: drag measurement	very low: drag compensation
control	none	angular	angular + linear
dynamic range	space fixed parallel transport	local level along- cross- radial	one component
proof masses	free floating = SST	constrained = gradiometry	
free floating: SST	hi-lo SST: GNSS + LEO	lo-lo SST; LEO cloud	lo-lo SST configuration (CART-WHEEL etc.)
constrained: gradiometry	differential accelerometry	rotating diff.-acc.	rotating torsional
arm length	hi-lo SST	lo-lo SST	inside S/C
number of gradiometric components	nine	full tensor	diagonal
temperature	ambient	high temperature superconducting	low temperature superconducting
			Eötvös
			one component

3.3. DYNAMIC RANGE

The relative precision also depends on the absolute signal range to be measured. It is the so-called dynamic range that has to be covered by the measuring device. There are essentially three options available that affect the necessary dynamic range.

The required range depends on the orientation of the gradiometer instrument during its transport along the orbit. If it is transported parallelly (space stable) the full gradiometric signal range, from about +3000 E to -1500 E, has to be measured. If it is kept earth-oriented, with one axis along track, one cross-track and one radial, a DC value of -1500 E affects the diagonal along-track and cross-track component and one of 3000 E the radial component. The DC value is only high for the diagonal components but it is about two orders of magnitude smaller for the off-diagonal components or for the difference $\Gamma_{xx} - \Gamma_{yy}$. The signal variation about the DC-value, i.e. the AC-range, is approximately the same for all components.

One may also focus on a certain limited spectral band and optimise the instrument for this measurement bandwidth.

3.4. USE OF EFFECTS SUCH AS RESONANCES

The sensitivity of the gravity sensor can be increased by taking advantage of a relative signal amplification at or around a resonance frequency. This was the idea behind the rotating gradiometer by Forward (1974). With increased baseline lengths this leads to concepts of configuration flights of several satellites. The characteristic sensitivities for the various options of configurations or multi-satellite missions need to be assessed.

From the above considerations one can conclude that a GRACE-like configuration, actively maintained at 200 km altitude, may be an interesting option for a future satellite gravity mission. A serious limitation resides in the fact that only the along track component of the gravity gradient tensor would be measured. Thus, sufficient decorrelation of the gravity signals will prove difficult and, in particular, aliasing poses a severe challenge, when trying to recover time variable gravity signals.

4. Gravity Field and Earth Sciences

We turn now to the initial question. Assuming GRACE and GOCE will meet their mission objectives is there still a scientific need for further satellite gravity missions? This is the central question to be addressed at this workshop. Let us try to add a few considerations to the discussion of the experts in the field.

In ESA (1999) an assessment of the scientific need of improved gravity field knowledge is given from the perspective of the GOCE mission. Thus, in particular the improvement of the quasi-static field for geophysics, oceanography, geodesy, glaciology and sea level research has been considered there. The results are summarized in Table II, taken from ESA (1999). The science rationale of the GRACE mission is based on a similar exercise, compare Committee on Earth Gravity from Space (1997). There the emphasis is on the time variable gravity field. The science issues are summarized in Table III.

Three applications of the gravity field in earth sciences have to be distinguished. First, any redistribution or transport of mass in the earth's system results in a change of the earth's gravity field. Thus, the measurement of these temporal variations will serve the analysis of transport processes. Second, the geoid (the equipotential surface at mean sea level) corresponds to the surface of a hypothetical ocean at rest. The quasi-stationary deviation of the actual ocean surface from the geoid, the dynamic topography, can be directly related to ocean surface circulation. Third, the differences between the actual geoid and the surface of a model earth in hydrostatic equilibrium are called geoid anomalies. Analogously the difference between actual gravity and that of such a model earth are denoted gravity anomalies. Both, geoid and gravity anomalies are a measure of the deviation of the actual earth from its state in hydrostatic equilibrium. They reflect the dynamic processes in the oceanic and continental lithosphere and in the upper mantle.

TABLE II
Static Gravity Field, Scientific Requirements

APPLICATION		ACCURACY		SPATIAL RESOLUTION half wavelength - D (km)
		Geoid (cm)	Gravity (mgal)	
SOLID EARTH	lithosphere / upper mantle density		1-2	100
	continental lithosphere rifts		1-2	50-100
	tectonic motions		1-2	20-100
	seismic hazards		1	100
ocean lithosphere / asthenosphere			0.5	100-200
OCEANOGRAPHY	short scale	1-2		100
		0.2		200
	basin scale	~ 0.1		1000
ICE SHEETS	rock basement		1-5	50-100
	ice vertical movements	2		100-1000
GEODESY	levelling by GPS	1		100-1000
	unified height systems	1		100-20000
	INS		~ 1-5	100-1000
	orbits		~ 1-3	100-1000
SEA LEVEL CHANGE		Many of the above applications, with their specific requirements, are relevant to studies of sea level change		

TABLE III
Time Variable Gravity Field, Scientific Requirements

GEODYNAMIC EFFECT	MAGNITUDE		SPATIAL RESOLUTION [KM]	MAIN PERIODS
	Geoid [mm]	Gravity [μ gal]		
Tides (oceans, solid earth)	100 – 150		50 – 5000	daily, semi-daily, semi-monthly
Atmosphere (IB, NIB, vertical integration)	15		200 – 2000	annual, seasonal, daily, others
Oceans (Sea level, currents)	10 – 15		100 – 2000	seasonal, secular
Hydrology (snow, rain, runoff, precipitation, evaporation, reservoirs, ground water)		10	10 – 1000	daily to annual
Postglacial rebound		10	1000 – 10000	secular
Polar ice & glaciers		5	100 – 1000	secular
Solid Earth				
• Earthquakes	0,5		10 – 100	single events
• Volcanism	0,5		10 – 100	single events
• Tectonics	?		>500	secular
• Core and Mantle	?		>5000	secular

4.1. GEOID AND GRAVITY ANOMALIES

Geoid anomalies reflect primarily mass anomalies of large to medium scale spatial extent; gravity anomalies are more related to medium and short scale features. Both, geoid and gravity anomalies, represent density contrasts in the lithosphere and upper mantle. The expectation is that after GOCE geoid anomalies are known with a precision of 1 cm and gravity anomalies with one of 0.1 mGal, both with a spatial resolution of about 100 km. With future satellite gravity field missions the spatial resolution can probably be increased to 50 km to 60 km. In addition, the global consistency, reliability and precision of the GOCE results could be improved further. This implies that future missions could, on the one hand, recover missing details, such as the gravity signals due to rift zones or smaller sedimentary basins, and on the other hand, increase the certainty and level of detail of already known gravity features. Very likely, any improvement in resolution beyond this cannot be measured from space. It would have to be deduced from airborne and terrestrial measurements.

At least as important as improvement of global gravity knowledge beyond that of GOCE is its adequate use in geophysical modelling. Earth sciences suffer from a profound lack of direct information about the physics and composition of the earth's interior. Everything is derived in an indirect manner. The three primary sources are, most importantly, the analysis of the propagation of seismic waves and, the earth's magnetic and gravity field. All further information is either even more indirect or not globally representative. These are, for example, surface topography, glacial isostatic adjustment, tectonic motion, earth rotation, the laboratory analysis of crust and mantle material and planetology. The trend is towards more and more comprehensive earth models, taking most or all of this information into account; global models or models tailored toward specific geophysical phenomena. Seismic tomography has brought a wealth of information about the dynamics of the earth interior. However, due to the nonlinear character of the seismic inversion problem and, in addition, due to the uneven distribution of seismic sources and seismic stations, some fundamental limitations remain, compare *e.g.* Snieder and Trampert (2000). Like seismic inversion, the estimation of the interior density structure from geoid or gravity anomalies is an inverse problem. Its character and its null space are very different, however, from that of seismic tomography. Therefore, the combination of the two can lead to significantly improved earth models. May be the greatest deficiency is currently the uncertainty in the rheological parameters. This aspect is discussed in King (2002). The interpretation of geoid anomalies has significantly improved in recent years. Examples are Lithgow-Bertelloni and Richards (1998), Simons and Hager (1997), or Vermeersen (2003). Also the joint inversion made progress, see *e.g.* Zerbini *et al.* (1992) or Negredo *et al.* (1999). Nevertheless, global integrated earth modelling in the above sense is still only in an early stage.

4.2. DYNAMIC OCEAN TOPOGRAPHY

Dynamic ocean topography is the quasi-stationary deviation of the actual ocean surface from the geoid (the hypothetical ocean surface at rest). It can be determined from the difference of satellite altimetry, measuring the actual ocean surface, and an accurate geoid model. It requires the time variable part to be eliminated (by means of averaging and using models). Ocean topography is only one to two meters in amplitude. It requires, both, the altimetric surface and the geoid to be determined with centimeter precision. When analysing future needs, after GRACE and GOCE, several questions have to be analysed:

- If the geoid could be derived with sub-centimeter precision, is it possible to determine the quasi-stationary ocean surface with this precision, too?
- What means quasi-stationary? At what temporal and spatial scales does ocean topography change? What are the shortest spatial scales at which a quasi-stationary dynamic topography can still exist? What is their relationship to the Rossby-radius?
- How to deal with the so-called omission part of the geoid, *i.e.* that (short scale) part of the geoid that cannot be sensed from space?

Almost certainly GOCE will not meet all requirements of oceanography in terms of spatial resolution and precision. An ultimate goal could be to determine the geoid with a precision of 1 cm and with a spatial resolution of 50 km.

A point of interest could be to translate the concept of direct measurement of dynamic topography to atmospheric modelling. In oceanography the dynamic topography determines the barotropic flow, see Open University Course Team (1989). The same concept holds for the atmosphere. Atmospheric sounding by GPS yields the geometry of the signal paths along with atmospheric pressure. Together with a global gravity field model the deviations between equipotential and equipressure surfaces could be derived, in principle.

4.3. TEMPORAL VARIATIONS OF THE GRAVITATIONAL FIELD

Gravity variations with time are the result of mass transport in atmosphere, oceans, hydrosphere, glacial areas and within the earth or of mass exchange between those components. Most of these processes are associated to global change phenomena and are essential for the establishment of mass balance within the earth system.

Furthermore, our understanding of the response of our planet to tidal forcing and the study of mass changes in the deep interior (inside the fluid outer core and at the core/mantle boundary) will greatly benefit from the measurement of time changes of the gravity field. Generally, temporal variations are very small; they occur in a wide range of time scales from sudden events, sub-daily, daily, seasonal, to long periodic and secular; they cover all spatial scales, too, from changes in flattening (see *e.g.* the recent work by Cox and Chao, 2002, and Cazenave and Nerem, 2002), via phenomena related to ocean areas or continental areas, only to

regional or local events such as those related to ground water variations of estuaria or to earthquakes, respectively.

The field of measurement of gravity time variations with satellites is largely open. GRACE is the first serious attempt in this direction, apart from the analysis of the time changes of the very low zonal harmonics from LAGEOS tracking data. The basic challenges are, (1) the small size of the effects to be measured, (2) the broad range of time scales to be covered, (3) the separation of the individual contributions, and (4) aliasing. A medium term strategy must therefore take into consideration the following elements:

- monitoring — a concept based on a series of missions,
- appropriate sampling in space and time — this may lead to a requirement of a formation flight of several satellites,
- complementary space missions and data sets needed for the separation of individual contributions and for comprehensive modelling.

References

- Cazenave, A. and Nerem, R. S.: 2002, 'Redistributing Earth's mass', *Science* **297**, 783–784.
- Colombo, O. L. and Chao, B. F.: 1993, 'Global Gravitational Change from Space in 2010', in H. Montag and Ch. Reigber (eds.), *Geodesy and Physics of the Earth, IAG-Symposia* **112**, Springer, Heidelberg, pp. 71–75.
- Colombo, G., Arnold, D. A., Binsack, J. H., Gay, R. H., Grossi, M. D., Lauman, D. A., Orringer, O.: 1976, 'Dumbell Gravity-Gradient Sensor: A New Application of Orbiting Long Tethers', *Reports on Geoastronomy* **2**, Smithsonian Astrophysical Observatory, Cambridge, Ma., 59 pp.
- Committee on Earth Gravity from Space: 1997, *Satellite Gravity and the Geosphere*, National Academy Press, Washington, D.C., 112 pp.
- Cox, C. M. and Chao, B. F.: 2002, 'Detection of a large scale mass redistribution in the terrestrial system since 1998', *Science* **297**, 831–833.
- Drinkwater, M. R., Floberghagen, R. F., Haagmans, R., Muzi, D., Popescu, A.: 2003, 'GOCE: ESA's First Earth Explorer Core Mission', *Space Sci. Rev.*, this volume.
- ESA: 1999, *Gravity Field and Steady-State Ocean Circulation Mission*, ESA SP-1233(1), ESA Publications Division, Noordwijk, 217 pp.
- Forward, R. L.: 1974, 'Review of Artificial Satellite Gravity Gradiometer Techniques for Geodesy', in G. Veis (ed.), *Proc. Intern. Symp. on the Use of Artificial Satellites for Geodesy and Geodynamics*, Publ. National Technical University, Athens, pp. 157–192.
- King, S. D.: 2002, 'Geoid and topography over subduction zones: The effect of phase transformations', *JGR* **107**, B1, ETG 2-1–2-10.
- Lithgow-Bertelloni, C., Richards, M. A.: 1998, 'The dynamics of cenozoic and mesozoic plate motions', *Reviews of Geophysics* **36**, 1, 27–78.
- Lumley, J. M., White, J. P., Barnes, G., Huang, D., Paik, H. J.: 2002, 'A Superconducting Gravity Gradiometer Tool for Exploration', unpublished manuscript.
- Negredo, A. M., Carminati, E., Barba, S., Sabadini, R.: 1999, 'Dynamic Modelling of Stress Accumulation in Central Italy', *Geophys. Res. Letters* **26**, 13, 1945–1948.
- Open University Course Team: 1989, G. Bearman (ed.), *Ocean Circulation*, Butterworth-Heinemann, Oxford, 238 pp.
- Paik, H. J. and Richard, J.-P.: 1986, *Development of a Sensitive Superconducting Gravity Gradiometer for Geological and Navigational Applications*, NASA Contractor Report 4011.

- Rummel, R.: 1986, 'Satellite Gradiometry', in H. Sünkel (ed.), *Mathematical and Numerical Techniques in Physical Geodesy, Lecture Notes in Earth Sciences 7*, Springer, Berlin, pp. 318–363.
- Rummel, R., and van Gelderen, M.: 1992, 'Spectral Analysis of the Full Gravity Tensor', *Geophys. J. Int.* **111**, 1, 159–169.
- Simons, M. and Hager, B. H.: 1997, 'Localization of the gravity field and the signature of glacial rebound', *Nature* **390**, 500–504.
- Snadden, M. J., McGuirk, J. M., Bayer, P., Haritos, K. G., Kasevich, M. A.: 1998, 'Measurement of the Earth's Gravity Gradient with an Atom Interferometer-Based Gravity Gradiometer', *Physical Review Letters* **81**, 5, 971–974.
- Snieder, R. and Trampert, J.: 2000, 'Linear and nonlinear inverse problems', in A. Dermanis, A. Grün, and F. Sanso (eds.), *Geomatic Methods for the Analysis of Data in the Earth Sciences, Lecture Notes in Earth Sciences 95*, Springer, Berlin, pp. 93–164.
- Touboul, P.: 2003, 'Microscope instrument development, lessons for GOCE', *Space Sci. Rev.*, this volume.
- Vermeersen, B.: 2003, 'The potential of GOCE in constraining the structure of the crust and lithosphere from post-glacial rebound', *Space Sci. Rev.*, this volume.
- Wells, W. C.,(ed.): 1984, *Spaceborne Gravity Gradiometers*, NASA Conference Publications, Scientific and Technical Information Branch, pp. 86.
- Williamstown Report: 1969, *The Terrestrial Environment. Solid-Earth and Ocean Physics, Applications of Space and Astronomic Techniques*, Report of a study at Williamstown, Mass. NASA.
- Zarembinski, S.: 2003, 'On superconductive gravity gradiometry in Space ', *Space Sci. Rev.*, this volume.
- Zerbini, S., Achache, J., Anderson, A. J., Arnet, F., Geiger, A., Klingelé, E., Sabadini, R., Tinti, S.: 1992, *Study of Geophysical Impact of High-Resolution Earth Potential Field Information*, ESA-study, final report, pp. 282.

I: PRECISE ORBIT DETERMINATION AND GRAVITY FIELD MODELLING

STRATEGIES FOR PRECISE ORBIT DETERMINATION OF LOW EARTH ORBITERS USING THE GPS

U. HUGENTOBLER and G. BEUTLER

Astronomical Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

Received: 9 August 2002; Accepted in final form: 4 November 2002

Abstract. Considerable experience accumulated during the past decade in strategies for processing GPS data from ground-based geodetic receivers. First experience on the use of GPS observations from spaceborne receivers for orbit determination of satellites on low altitude orbits was gained with the launch of TOPEX/POSEIDON ten years ago. The launch of the CHAMP satellite in July 2000 stimulated a number of activities worldwide on improving the strategies and algorithms for orbit determination for Low Earth Orbiters (LEOs) using the GPS.

Similar strategies as for ground-based receivers are applied to data from spaceborne GPS receivers to determine high precision orbits. Zero- and double-differencing techniques are applied to obtain kinematic and/or reduced-dynamic orbits with an accuracy which is today at the decimeter level. Further developments in modeling and processing strategies will continuously improve the quality of GPS-derived LEO orbits in the near future. A significant improvement can be expected from fixing double-difference phase ambiguities to integer numbers. Particular studies focus on the impact of a combined processing of LEO and GPS orbits on the quality of orbits and the reference frame realization.

1. Introduction

The launch of the CHAMP satellite carrying on board one of the first reliable dual-band spaceborne GPS receivers and the fascinating missions such as GRACE and the upcoming GOCE mission for gravity field recovery, triggered a number of activities to use the Global Positioning System (GPS) Satellite-Satellite Tracking (SST) data for Precise Orbit Determination (POD) of Low Earth Orbiters (LEOs). The continuous data acquisition in combination with the accuracy of the phase observations is an important advantage with respect to other tracking systems. Table I contains names of the relevant scientific missions carrying spaceborne GPS receivers.

Considerable experience related to strategies for processing GPS data from ground-based receivers and orbit determination of GPS satellites was gained during the past decade. The International GPS Service (IGS) must be mentioned in this context: The IGS coordinates data acquisition, collection, and processing of a global network of more than 300 ground-stations, and provides high accuracy GPS orbits and GPS clock corrections on a routine basis.

Similar strategies as for ground-based receivers may be applied to the processing of data from orbiting GPS receivers with the goal to determine either kinematic



TABLE I
Selection of satellites carrying GPS receivers

Satellite	Launch	Mission
TOPEX/POSEIDON	Aug. 10, 1992	Altimetry
GPS/MET	Apr. 9, 1995	Atmospheric sounding
CHAMP	Jul. 15, 2000	Gravity field, magnetic field
SAC-C	Nov. 23, 2000	Earth observation
JASON-1	Dec. 7, 2001	Altimetry
GRACE	Mar. 17, 2002	Gravity field
ICESat	Dec. 15, 2002	Laser altimetry
GOCE	2006	Gravity field

trajectories or dynamic and reduced-dynamic orbits for the fast moving satellites. Observations may be processed in the zero-difference or the double-difference mode with baselines relating the satellite of interest with the ground-stations. Most commonly, the orbits of the GPS satellites are introduced as known, e.g., by taking the information from the IGS. Alternatively LEO and GPS orbits might be computed in the same program run to enforce a consistent treatment.

2. Principles of GPS Data Processing

Geodetic receivers register pseudorange (code) and phase measurements on both frequencies emitted by the GPS satellites (1575.42 MHz, $\lambda_1=19.0$ cm and 1227.60 MHz, $\lambda_2=24.4$ cm). The measurement precision of the two observables is of the order of one meter for the code and of a few millimeters for the phase. It is in particular the high precision of the second measurement type which offers a big variety of geodetic and navigation applications.

The observation equations governing the GPS measurements refer either to the pseudorange or to the phase. The code observation equation for receiver i and satellite k reads as

$$P_i^k = |\mathbf{r}_i - \mathbf{r}^k| + c\Delta t_i - c\Delta t^k + T_i^k + I_i^k + \varepsilon_i^k \quad (1)$$

where \mathbf{r}_i and \mathbf{r}^k denote the position vectors of the receiving antenna at the observation epoch (signal reception time) and the satellite antenna at the observation epoch minus the light travel time, respectively, Δt_i and Δt^k represent synchronization parameters for receiver and satellite clocks to GPS time, T_i^k and I_i^k denote the corrections due to the tropospheric and the ionospheric refraction, and ε_i^k stands for the observation error.

The phase observation equation

$$L_i^k = |\mathbf{r}_i - \mathbf{r}^k| + c\Delta t_i - c\Delta t^k + T_i^k - I_i^k + \lambda N_i^k + \varepsilon_i^k \quad (2)$$

is similar to the code observation equation except for the sign of the ionospheric refraction term and, more important, the appearance of an integer multiple of wavelengths λN_i^k . This *phase ambiguity term* points to a fundamental difference between the two measurement types. Phase measurements between a satellite and a receiver are known only up to the constant λN_i^k during a satellite pass which stays at the same value, as long as the receiver does not loose lock of the satellite. These phase ambiguities have to be estimated within the parameter estimation procedure. They constitute, apart from epoch-specific parameters such as clock corrections or kinematic coordinates, by far the largest set of unknowns.

The ionosphere-free linear combination of the observations at the two frequencies may be used to eliminate the dispersive effect of the ionosphere, an aspect which is also relevant for LEOs. The so-called geometry-free linear combination, on the other hand, removes all terms which are the same on both frequencies which allows it to study the impact of the ionosphere on the measurements. Other linear combinations may be used to identify the integer values of the phase ambiguities as long as they conserve the integer nature of the ambiguities. The so-called wide-lane linear combination with an effective wavelength of 86 cm is helpful in this context. The so-called Melbourne-Wübbena linear combination of the equations for code and phase removes the ionosphere as well as the geometry terms and allows it to find the ambiguity parameters without any knowledge of the geometry. The drawback is the noise of the combination as it is governed by the precision of the code observable. For more details see, e.g., (Hofmann-Wellenhof et al., 1997).

Differences between observations may be used to eliminate biases which are satellite- or receiver-specific such as clock corrections, by forming single- or double differences. Differences of observations from one receiver to different GPS satellites remove the receiver clock biases while differences referring to one satellite and different receivers (baselines) remove in essential the corresponding biases for the satellite. Differences (the so-called double differences) between pairs of receivers and satellites eliminate all clock corrections.

The creation of differences between observations is allowed as long as the eliminated biases can be assumed to stay constant during signal emission and reception times (due to receiver synchronization and light propagation time differences). GPS receivers, in fact, are built in such a way that measurements are acquired within a millisecond to an integer second. The concept of forming zero- and double differences of the original observations — commonly used for GPS data processing — is illustrated in Figure 1 for a LEO.

Differences may also be formed between successive observation epochs. In this way the phase ambiguities are eliminated. Procedures based on time differences are used for efficient data screening and for LEO orbit determination, see, e.g., (Bisnath and Langley, 1999) and (Bock et al., 2002).

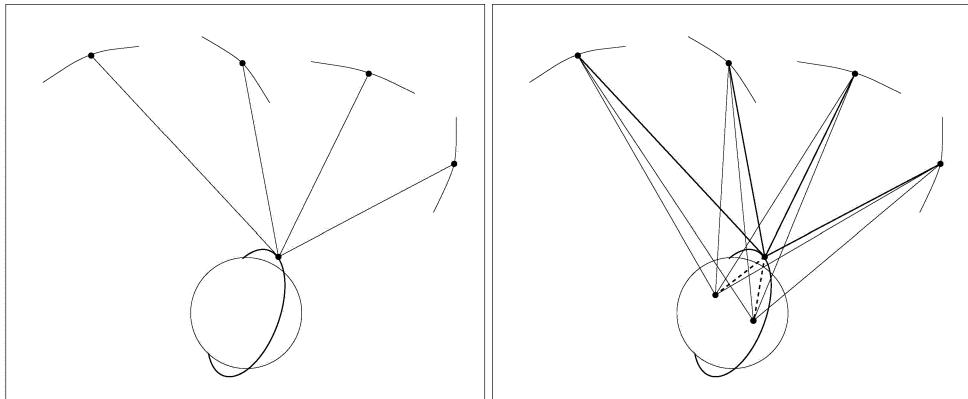


Figure 1. Schematic view of observations required to form zero-differences (*left*) and double-differences (*right*) for orbit determination of a LEO

Due to the correlation with clock parameters, the integer nature of the ambiguities can only be exploited on the double difference level. Various methods may be applied to find the integer values of the ambiguity parameters on both frequencies, see, e.g., (Theunissen, 2001). Fixing the ambiguities on the double difference level reduces the number of parameters and stabilizes the solution.

3. GPS Orbits and Reference Frame

For the computation of a precise LEO orbit using GPS spaceborne observations the orbits of the high orbiting GPS satellites are usually introduced as known. High precision orbits for GPS satellites are made available by the IGS, see, e.g., (Beutler et al., 1999). The orbits are computed by several IGS Analysis Centers which base their solutions on different subsets of the global IGS network of about 300 tracking stations. The IGS provides Earth orientation parameters, station coordinates and velocities, GPS clock corrections, troposphere zenith delays, and global maps of ionosphere electron content in addition to precise GPS orbits.

The IGS orbits are the result of a combination of the orbits stemming from the different IGS Analysis Centers which use different software packages and processing strategies. The final IGS orbits show an internal consistency of the individual solutions produced by the IGS Analysis Centers of 2–6 cm RMS per coordinate. An external accuracy of about 5 cm is confirmed by Satellite Laser Ranging (SLR) measurements.

The IGS satellite orbits are provided in the ITRF reference system which is realized by applying a no net rotation condition to the (about 50) ITRF core sites. The ITRF reference frame is ‘inherited’ by the LEO orbit determined from zero-difference GPS observations. In case of double-difference data processing the GPS orbits and the station coordinates must refer to the same reference system.

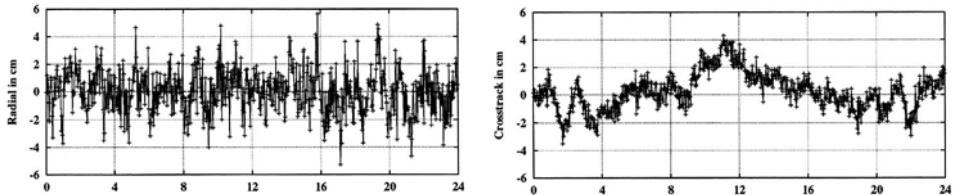


Figure 2. Position differences between kinematic orbit determined using undifferenced observations and true orbit in radial (*left*) and cross-track (*right*) direction (simulated data)

4. LEO Orbit Determination

Apart from the determination of a dynamic or reduced-dynamic orbit the data stemming from a spaceborne GPS receiver allow it to reconstruct a purely kinematic trajectory which does not make use of the equations of motion. Using the observations of all GPS satellites in view an independent position is computed for each observation epoch.

Zero-difference processing is very efficient — as long as the orbits of the GPS satellites are introduced as known — because no ground station data are required. The LEO orbit depends, however, implicitly on the ground station information through the GPS orbits and clock corrections computed in a separate computation or taken from an external source such as the IGS. The method requires the estimation of (real valued) phase ambiguities and of epoch-wise clock corrections for the receiver. For a typical LEO about 400 ambiguity parameters accumulate per day. In a least squares adjustment the clock corrections and the kinematic positions may be pre-eliminated and back-substituted in a successive processing run in order to keep the number of parameters manageable.

Figure 2 shows the residuals in radial and cross-track directions of kinematic positions determined using undifferenced simulated observations with respect to the true orbit. A noise of 3 mm on both carriers was assumed for simulating the phase observations and no ambiguities were introduced in addition to the initial ambiguities. Exactly as for Earth-fixed stations the noise in the radial (zenith) direction is about a factor of three larger than for the two other components. This is due to the fact that GPS satellites are observed only in the hemisphere pointing away from the Earth. The systematic variations of the residuals in the cross-track direction is due to the correlation of ambiguity parameters with clock parameters and coordinates (a similar behaviour is observed in the along-track component).

For double-difference processing the data from ground stations have to be used even if the GPS orbits are introduced as known in order to form baselines between the LEO and the ground stations. If more than one LEO could be analyzed together, baselines between the LEOs might be introduced as well.

Figure 3 shows an example for the baselines which are established at a particular epoch between the satellite CHAMP and 64 globally distributed ground stations. The conditions imposed were a minimum of three common GPS satellites observed

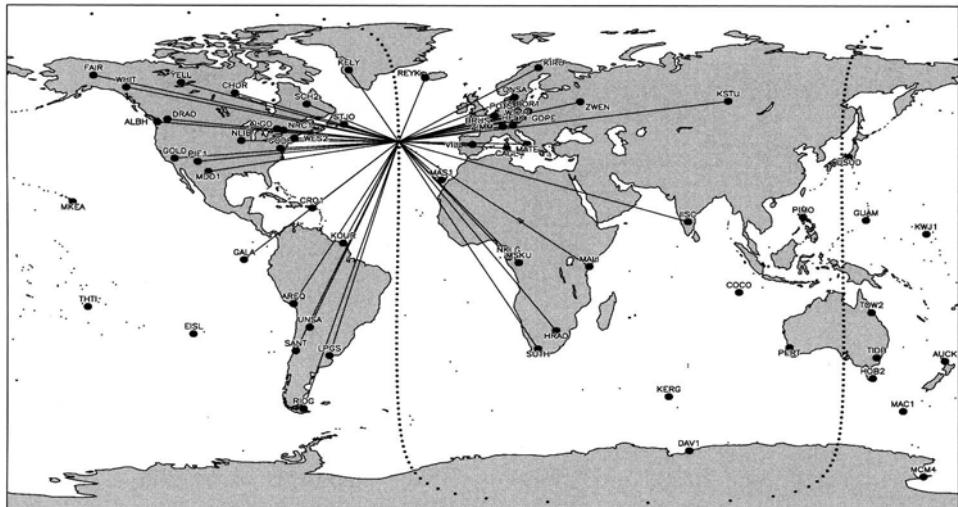


Figure 3. Baselines between LEO and ground stations

by both receivers to get redundancy, a positive elevation angle of GPS satellites observed by the LEO antenna, and an elevation angle larger than ten degrees for the ground station antennas. With these conditions baselines may be as long as 10'000 km and they are active for about 50% of the time on average.

No clock corrections need to be computed for LEO and for GPS satellites. On the other hand, coordinates and troposphere zenith delay parameters for the ground stations have to be estimated or introduced as known, e.g., on values provided by the IGS. The number of ambiguities is extremely large due to the rapid motion of the LEO and the associated rapid change of the configuration of the baselines. For 60 ground-stations more than 10'000 ambiguity parameters have to be handled per day.

5. Ambiguity Resolution to Integer Values

Due to the big number of ambiguity parameters required in the double-difference processing of LEO observations, ambiguity resolution implies long processing times on state of the art computers. The ambiguity parameters cannot be pre-eliminated, and an iterative solution scheme is necessary as described, e.g., in (Theunissen, 2001). Fixing the best determined ambiguities improves the probability for finding the correct integer values for other ambiguity parameters.

Tests with simulated and real data from the CHAMP satellite were carried out by (Švehla and Rothacher, 2002). The Melbourne-Wübbena linear combination of code and phase may be used as for Earth-fixed baselines to find in a first step integer values for the wide-lane linear combination. In a second step the integer values for the so-called narrow lane linear combination are fixed. The ability of

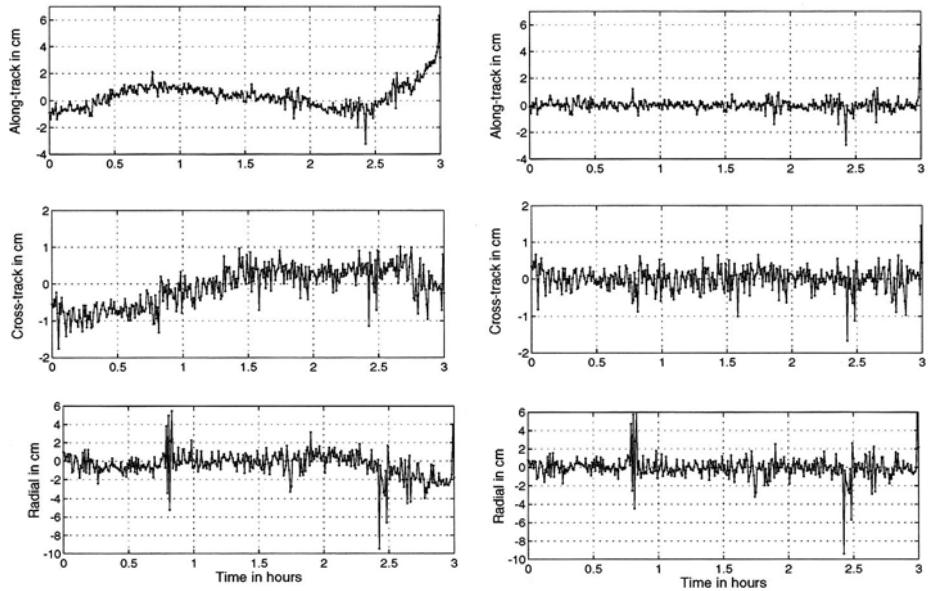


Figure 4. Position differences between kinematic orbit with float ambiguities (left) and fixed ambiguities (right) and true orbit (simulated data), from (Švehla and Rothacher, 2002)

the algorithm to identify the correct integer values for the wide-lane ambiguities depends critically on the accuracy of the pseudorange measurements, i.e., on the receiver performance.

Švehla and Rothacher simulated observations with a phase noise of 1 mm for the CHAMP satellite and for 105 globally distributed ground stations. Residuals of the kinematic positions for the double-difference solution with float ambiguities are given for the three components on the left-hand side of Figure 4. The corresponding values for the solution with the ambiguities fixed to their integer values using the wide-lane/narrow-lane approach are given on the right-hand side of the same figure. The systematic excursions of the residuals almost completely disappear when the ambiguities are fixed. Results of this kind are the motivation to develop operational strategies for ambiguity resolution for baselines involving a spaceborne receiver.

In tests based on 3 hours of real data from CHAMP together with 140 IGS ground stations Švehla and Rothacher were able to find the integer values for 42% of the wide-lane ambiguities and 27% of the narrow-lane ambiguities. The impact on the kinematic positions of the satellite is of the order of 5 cm.

6. Combined Processing of GPS and LEO

The strategies in the previous sections were all discussed under the premise that the orbits of the GPS satellites are introduced as a priori known. In this two-step approach the GPS orbits are computed independently of the determination of the

TABLE II
Topex orbit comparison with JPL's orbit (cm), from (Rim et al., 1995)

Strategy	Radial	Along-track	Cross-track	RSS
Two step	2.03	5.17	2.71	6.18
Single step A	1.73	4.45	3.26	5.78
Single step B	1.73	4.18	2.57	5.20

LEO orbit. Maximum advantage may, however, be made of the complete set of GPS observations by estimating orbits for GPS satellites *and* LEO in a common parameter estimation procedure. This *single-step approach* would guarantee a fully consistent treatment of ground station and LEO tracking data.

The correct use of the covariance information of the GPS orbits — instead of introducing them as fixed — may improve the LEO orbit, and the additional constraints on the high orbits might help to improve the GPS orbits. One could also expect an improvement of the GPS estimates for Earth orientation parameters or for geocenter coordinates by adding LEOs into the processing. The modeling of the LEO orbit must be sufficiently accurate, however. Otherwise biases of the low orbit might degrade the accuracy of the GPS orbits. In view of the high degree of freedom of a combined solution the interpretation of possible modeling problems could be very difficult.

Already half a decade ago (Rim et al., 1995) performed tests with single-step and two-step solutions of GPS derived precise orbits for TOPEX/POSEIDON. Some results of this experiment are reproduced in Table II. It lists the rms difference in the three components as well as the 3-dim rms of the TOPEX orbits as compared to the orbits generated independently at JPL. The reduced-dynamic TOPEX orbits are based on a double-difference approach, where solution A uses only baselines between the LEO and ground stations, while solution B includes, in addition, baselines between ground stations. An improvement in radial and along-track directions can be observed. Currently, however, the Center for Space Research (CSR), University of Texas at Austin, is applying a two-step approach for the determination of precise orbits for CHAMP because with today's accuracy of the IGS orbits no improvement is expected for the orbit accuracy of the satellite.

In a study on POD for the GOCE satellite (Visser and van den IJssel, 2001), no significant improvement of neither the LEO nor the GPS orbits could be found, but the danger of the propagation of LEO orbit modeling errors into the GPS orbits was identified. Tests carried out by (Zhu et al., 2002) using real data stemming from the CHAMP satellite indicate, on the other hand, an improvement of the reference frame realization as well as of the quality of the LEO and GPS orbits.

A dedicated analysis using simulated data from 64 globally distributed stations shows a small improvement of the formal a priori error of the GPS satellite po-

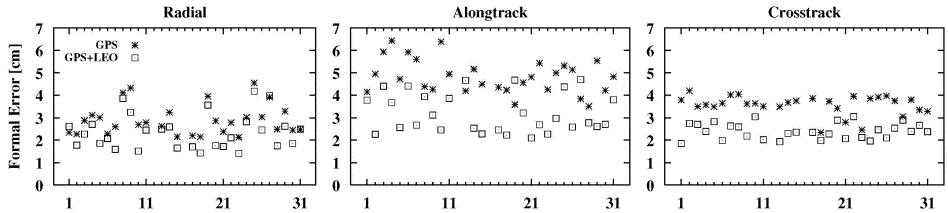


Figure 5. Formal errors a priori from a covariance simulation for the GPS satellite positions as a function of PRN number, processing of GPS only (stars) and of GPS and LEO together (boxes)

sitions. The results are based on double differenced observations using baselines between the LEO and ground stations. The orbit of the LEO is modeled dynamically, a scenario promising the biggest impact on the formal accuracy of the GPS orbits. Systematic errors were not considered. The results for the two-step and the single-step approach for all GPS satellites are provided in Figure 5. The two approaches lead to similar results in the radial direction while in the along-track and the cross-track directions small but significant improvements of the formal accuracy are observed for the combined processing strategy. This indicates an improved definition of the reference frame. The findings are supported by an improvement of the formal accuracy of the Earth's pole coordinates. The third component of the Earth orientation, UT1 resp. length of day, shows, however, a slight deterioration of the formal error. Because the GPS orbits are introduced as known for the two-step approach, the simulation does not allow it to interpret differences in the formal errors for the LEO orbit.

The same simulation also indicates a significant influence of the systematic errors of the LEO orbit on the GPS orbits via the correlation with the ambiguity parameters. Further studies are required to quantify the effects of a combined processing on the accuracy of the LEO orbit, the GPS orbits, and of the Earth orientation parameters.

7. Conclusions

The same GPS data processing strategies as those used to analyze observations from Earth-fixed stations may in principle be applied to spaceborne GPS receivers. The zero-, double-, and triple-difference approaches may be used to derive kinematic and reduced-dynamic LEO orbits. The accuracies of these orbits are estimated to be at or below the decimeter level for CHAMP, which has been verified by comparison with independent SLR observations (Boomkamp, 2002). It may be foreseen that ambiguity resolution will lead to a significant improvement of the kinematic orbits.

Whether the two-step approach (introducing the GPS orbits as known) or the single step approach (using the LEO as an additional station in a global network)

is the better strategy is an open question. In view of the high quality achieved by the IGS it seems rather difficult to come up with significant improvements in the (potentially time consuming) single step approach.

Acknowledgements

We wish to thank R. König and S. Y. Zhu, GeoForschungsZentrum Potsdam, S. Loyer, Centre National d' Etudes Spaciales, Toulouse, H. J. Rim, Center for Space Research, University of Texas at Austin, M. Rothacher and D. Švehla, Institute for Astronomical and Physical Geodesy, Technical University Munich, P. Visser, Delft Institute for Earth-Oriented Space Research, Delft Technical University, for helpful and interesting discussions.

References

- Beutler, G., M. Rothacher, S. Schaer, T. A. Springer, J. Kouba, R. E. Neilan: 1999, 'The International GPS Service (IGS): An Interdisciplinary Service in Support of Earth Sciences'. *Advances in Space Research* **23**, No. 4, 631–653.
- Bisnath, S. B., R. B. Langley: 1999, 'Precise A Posteriori Geometric Tracking of Low Earth Orbiters with GPS'. *Canadian Aeronautics and Space Journal* **45**, No. 3, 245–252.
- Bock, H., U. Hugentobler, T. A. Springer, G. Beutler: 2002, 'Efficient Precise Orbit Determination of LEO Satellites using GPS'. *Advances in Space Research* **30**, No. 2, 295–300.
- Boomkamp, H.: 2002, 'CHAMP Orbit Comparison Campaign'. <http://nng.esoc.esa.de/gps/campaign.html>
- Bertiger, W. J., Y. E. Bar-Sever, E. J. Christensen, E. S. Davis, J. R. Guinn, B. J. Haines, R. W. Ibanez-Meyer, J. R. Jee, S. M. Lichten, W. G. Melbourne, R. J. Muellerschoen, T. N. Munson, Y. Vigue, S. C. Wu, T. P. Yunck, B. E. Schutz, P. A. M. Abusali, H. J. Rim, M. M. Watkins, P. Willis: 1994, 'GPS Precise Tracking of TOPEX/POSEIDON: Results and Implications'. *Journal of Geophysical Research* **99**, No. C12, 24,449–24,464.
- Hofmann-Wellenhof, B., H. Lichtenegger, J. Collins: 1995, 'GPS, Theory and Practice', fourth edition, Springer, Wien, New York, 1997.
- Rim, H. J., B. E. Schutz, P. A. M. Abusali, B. D. Tapley: 1995, 'Effect of GPS Orbit Accuracy on GPS-Determined TOPEX/POSEIDON Orbit'. In *Proceedings of ION GPS-95*, 613–617, September 12–15, 1995.
- Švehla, D., M. Rothacher: 2001, 'Kinematic Orbit Determination of LEOs Based on Zero or Double-Difference Algorithms Using Simulated and Real SST GPS Data'. In *Vistas for Geodesy in the New Millennium*, Schwarz K.P, Adam J., (Eds), Springer IAG Vol. 125, 322–328.
- Theunissen, P. J. G: 2001, 'The Probability Distribution of the Ambiguity Bootstrapped GPS Baseline. *Journal of Geodesy* **75**, 267–275.
- Visser, P. N. A. M., J. van den IJssel: 2001, 'GPS-Based Precise Orbit Determination of the Very Low Earth-Orbiting Gravity Mission GOCE'. *Journal of Geodesy* **74**, 590–602.
- Zhu, S., H. Neumayer, F. H. Massmann, C. Shi, C. Reigber: 2002, 'Impact of Different Data Combinations on the CHAMP Orbit Determination'. In *Proceedings of the IGS Workshop 2002*, in press, Ottawa, April 8–12, 2002.

Address for Offprints: U. Hugentobler, Astronomical Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland, hugentobler@aiub.unibe.ch

AIMING AT A 1-CM ORBIT FOR LOW EARTH ORBITERS: REDUCED-DYNAMIC AND KINEMATIC PRECISE ORBIT DETERMINATION

P.N.A.M. VISSER (pieter.visser@lr.tudelft.nl) and

J. VAN DEN IJSSEL (jose.vandenijssel@lr.tudelft.nl)

Delft Institute for Earth-Oriented Space Research, Delft University of Technology

Received: 2 May 2002; Accepted in final form: 25 September 2002

Abstract. The computation of high-accuracy orbits is a prerequisite for the success of Low Earth Orbiter (LEO) missions such as CHAMP, GRACE and GOCE. The mission objectives of these satellites cannot be reached without computing orbits with an accuracy at the few cm level. Such a level of accuracy might be achieved with the techniques of reduced-dynamic and kinematic precise orbit determination (POD) assuming continuous Satellite-to-Satellite Tracking (SST) by the Global Positioning System (GPS). Both techniques have reached a high level of maturity and have been successfully applied to missions in the past, for example to TOPEX/POSEIDON (T/P), leading to (sub-)decimeter orbit accuracy. New LEO gravity missions are (to be) equipped with advanced GPS receivers promising to provide very high quality SST observations thereby opening the possibility for computing cm-level accuracy orbits. The computation of orbits at this accuracy level does not only require high-quality GPS receivers, but also advanced and demanding observation preprocessing and correction algorithms. Moreover, sophisticated parameter estimation schemes need to be adapted and extended to allow the computation of such orbits. Finally, reliable methods need to be employed for assessing the orbit quality and providing feedback to the different processing steps in the orbit computation process.

Keywords: precise orbit determination, reduced-dynamic, kinematic, GPS, LEO

1. Introduction

The launch of CHAMP in July 2000 has triggered significant efforts by many scientific institutes in the field of precise orbit determination (POD). Without very high precision orbit determination, one of the most important mission objectives of CHAMP cannot be reached, namely a significant improvement in global Earth gravity field modeling (Reigber et al., 1999). High-precision orbit determination becomes even more of a challenge for the upcoming GRACE mission (launch in March 2002) and the future GOCE mission (expected launch in early 2006). These missions are much more demanding in terms of gravity field modeling performance than CHAMP and even more stringent orbit accuracy requirements are imposed. In order to get the most out of these missions, an orbit accuracy at the cm level is aimed at (NRC, 1997; ESA, 1999). All previously mentioned missions are Low Earth Orbiters (LEOs) flying at very low altitudes, in the 240–450 km height range



above the Earth's surface. An orbit accuracy level of a few cm can only be achieved with high-quality, continuous tracking, such as achieved by high-quality spaceborne receivers that acquire Satellite-to-Satellite Tracking (SST) observations to the Global Positioning System (GPS), appropriate data preprocessing and correction schemes, and orbit parameter estimation techniques. Two such techniques will be addressed below. They are referred to as reduced-dynamic and kinematic precise orbit determination.

POD is also in the spotlight due to the recent launches of Jason-1 (December 2001), equipped with a GPS receiver, and ENVISAT (March 2002), equipped with a DORIS receiver. For these missions, an orbit accuracy of about 1 cm is aimed at as well (particularly for the radial direction), but in these cases the satellites fly at higher altitudes, 1336 and 800 km, respectively, and a high orbit precision is required to support the interpretation of radar altimeter observations (Jason-1 and ENVISAT) and images taken by the Synthetic Aperture Radar (ENVISAT).

After introducing the principles of kinematic and reduced-dynamic POD, a few recent results will be discussed, including simulation studies to assess achievable orbit accuracies for GOCE and to identify critical items. After this, attention will be paid to activities related to the CHAMP mission, which is now in the center of LEO POD activities. Moreover, POD of CHAMP is the focus of an international initiative by the International GPS Service (IGS) to improve LEO POD. This initiative includes the application and tuning of reduced-dynamic and kinematic orbit determination schemes and upgrading of observation preprocessing and correction algorithms.

2. Precise Orbit Determination

In general, orbit determination is composed of the following steps: (1) collecting, preprocessing and correcting tracking observations, (2) defining standards and reference systems, (3) defining dynamic (optional) and satellite models, (4) estimating parameters and (5) assessing/verifying the orbit accuracy.

The focus in the remainder of this paper will be on the 4th step, namely two parameter estimation techniques referred to as reduced-dynamic and kinematic orbit determination. In dynamic orbit determination, the orbit is obtained by determining those state vector values, e.g., initial position and velocity at the start of an orbital period, atmospheric and solar radiation scaling parameters, etc., in such a way that the resulting orbit represents all observations best in a least squares sense. The orbit is completely determined by the dynamic model implemented in the equations of motion.

Reduced-dynamic orbit determination might be defined as a dynamic orbit determination augmented with an additional set of dynamic parameters, e.g., empirical accelerations. In general, at least two important parameters can be optimized (or manipulated) with respect to these accelerations: the correlation length or time

interval τ and the a priori standard deviation σ_{acc} (see also (Wu et al., 1990)). These parameters should reflect the quality of the a priori dynamic models (Visser and van den IJssel, 2000). Another important parameter is the weight of the tracking observations, which in fact has to be optimized simultaneously with the empirical acceleration parameters. Reduced-dynamic orbit determination may be seen as a trade-off between using a priori knowledge in the form of dynamic models and geometric information content of tracking observations. In the case of an optimal trade-off, reduced-dynamic orbit determination should give the best orbit solution possible using a certain data set of tracking observations. Different implementations of reduced-dynamic orbit determination techniques exist, which can be divided into sequential and batch parameter estimation methods, for example Kalman filtering and batch least squares estimation with constraint equations. Dynamic orbit determination can be considered to be a limit of reduced-dynamic orbit determination and is effectively obtained by specifying $\tau = \infty$ and $\sigma_{\text{acc}} = 0$.

Kinematic orbit determination is based on the idea that no a priori dynamic models are required in deriving the orbit positions of the LEO satellite. In a way, it can be regarded as the other limit of reduced-dynamic orbit determination, where the weight of the a priori dynamic model approaches the value zero by specifying $\tau = 0$ and $\sigma_{\text{acc}} = \infty$. Kinematic orbit determination can be categorized in point positioning methods, where the position of the LEO satellite is obtained for each epoch by geometric relations between the GPS observations and the GPS and LEO positions (Bock et al., 2001; Svehla and Rothacher, 2001), and sequential estimation methods, where use can be made of, e.g., a Kalman filter (Byun and Schutz, 2001).

In the case of kinematic POD, the point positioning method can only result in cm level orbits when using phase observations and fixing (a significant percentage of) the ambiguities. The sequential estimation method is more flexible with respect to the form in which the observations are used and the character and amount of unknowns (Bisnath and Langley, 2001).

The reduced-dynamic orbit determination technique allows the inclusion of dynamic models. This technique also allows the estimation of certain dynamic parameters, including the initial position and velocity of LEO and GPS satellites and empirical accelerations. An important issue that needs to be addressed concerns the computation of the orbits of the GPS satellites: it has to be assessed whether the best LEO orbit is obtained by estimating this orbit simultaneously with the LEO orbit or not.

3. Results

The capability of computing sub-decimeter accuracy orbits has been demonstrated for a few satellite missions in the past. Further improvements in terms of orbit accuracy, down to the cm level, are expected to be realized for currently flying

satellites. The following two sections will address results obtained with a few typical satellites that are equipped with a space-borne GPS receiver.

3.1. PAST MISSIONS AND FEASIBILITY STUDIES

Decimeter level accuracy orbits were for the first time demonstrated for the T/P satellite, which was not only equipped with a GPS receiver, but also with a DORIS receiver and an array of SLR retro-reflectors allowing an independent assessment of the achieved orbit quality (Smith et al., 1994). The accuracy in the radial direction is estimated to be around 3 cm (Tapley et al., 1994). Although the GPS receiver allowed tracking of at most six GPS satellites simultaneously, the concepts of reduced-dynamic and kinematic POD could be tested successfully for this satellite.

The reduced-dynamic POD technique should be relatively insensitive to dynamic model errors provided that a proper accuracy assessment of the dynamic models is made and in conjunction the empirical acceleration unknowns are properly tuned. Reduced-dynamic and dynamic T/P orbits computed with the JGM2 gravity field model display similar differences as dynamic orbits computed with the JGM2 and JGM3 gravity field models (Figure 1), where the JGM3 gravity field model is a significant improvement over JGM2 (Tapley et al., 1994). In other words, the reduced-dynamic POD technique in combination with a relatively inaccurate gravity field model provides an orbit with about the same accuracy as a dynamic orbit using a more accurate gravity field model, thereby proving the feasibility of this technique.

The reduced-dynamic T/P orbit accuracy clearly depends on the correlation times and a priori and steady-state standard deviations (denoted by σ_p and σ_0). Different settings of these parameters may lead to orbit differences of the order of 10 cm 3-dimensionally for this relatively high-flying satellite (Figure 2).

Figure 3 displays the orbit differences between a T/P orbit obtained by a sequential filter kinematic technique and a high-precision reduced-dynamic orbit solution. The tracking data consisted of undifferenced single-frequency pseudo-range and phase observations. This experiment suggest that a 3-dimensional orbit accuracy of about 30 cm is possible for T/P using a kinematic POD approach. Other experiments have indicated that this can be improved to better than 15 cm, *cf.* (Byun and Schutz, 2001).

Probably the most challenging planned satellite from the viewpoint of POD will be the GOCE satellite, which will fly at an extremely low altitude around 240–250 km. A detailed orbit accuracy assessment has been conducted by full-scale reduced-dynamic and kinematic POD simulations. It was found that the eventually achievable orbit accuracy depends on the success of carrier phase ambiguity fixing (especially for the kinematic approaches) and thus on the quality of the GPS receiver (especially the quality of the pseudo-range observations). When ambiguity fixing cannot be done, the accuracy of the reduced-dynamic orbits depends to some

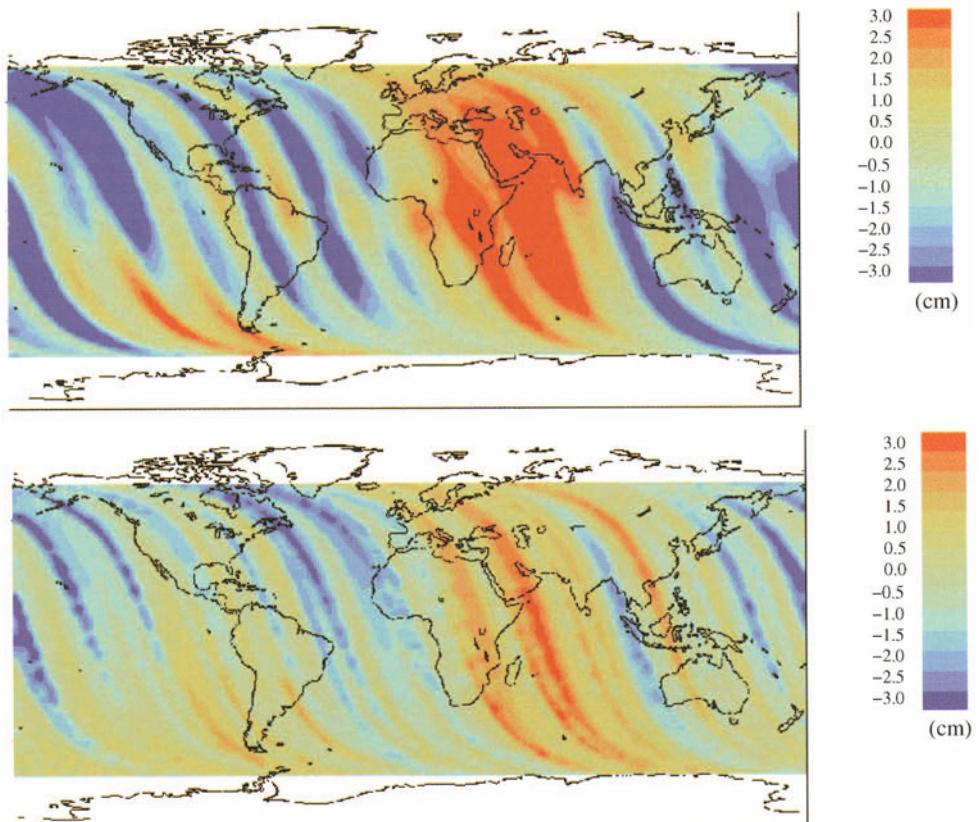


Figure 1. Comparison between reduced-dynamic and dynamic T/P orbits using different gravity field models (orbits by *F. Nouel (CNES, France)*, Kalman filter). In this particular case the orbits were based on DORIS tracking data, which is also near-continuous in time. Similar results were obtained with GPS data (Smith et al., 1994). The top figure displays the radial orbit differences between dynamic orbits obtained using JGM2 and JGM3. The bottom figure displays radial orbit differences between a dynamic orbit and a reduced dynamic orbit obtained using JGM2.

extent on the quality of the a priori dynamic models and the risk arises of “biting one’s own tail”, because the primary mission objective of GOCE is to obtain a very accurate and high-resolution gravity field model. The GOCE simulations indicate that the achievable 3-dimensional orbit accuracy is about 3 cm when the ambiguities can be fixed (reduced-dynamic and kinematic), but this accuracy can deteriorate significantly when the ambiguities cannot be fixed (see Tables 4 and 5 in (Visser and van den IJssel, 2000)). The results were inconclusive concerning the issue of simultaneously estimating the LEO and GPS orbits or fixing the orbits of the latter, e.g., to the solution provided by the IGS. Similar GOCE orbit accuracies were obtained with both concepts. The simulations included the incorporation of accelerometer observations in the POD. It was assumed, however, that these observations must meet the requirements of the mission (“specs”) and the issue of

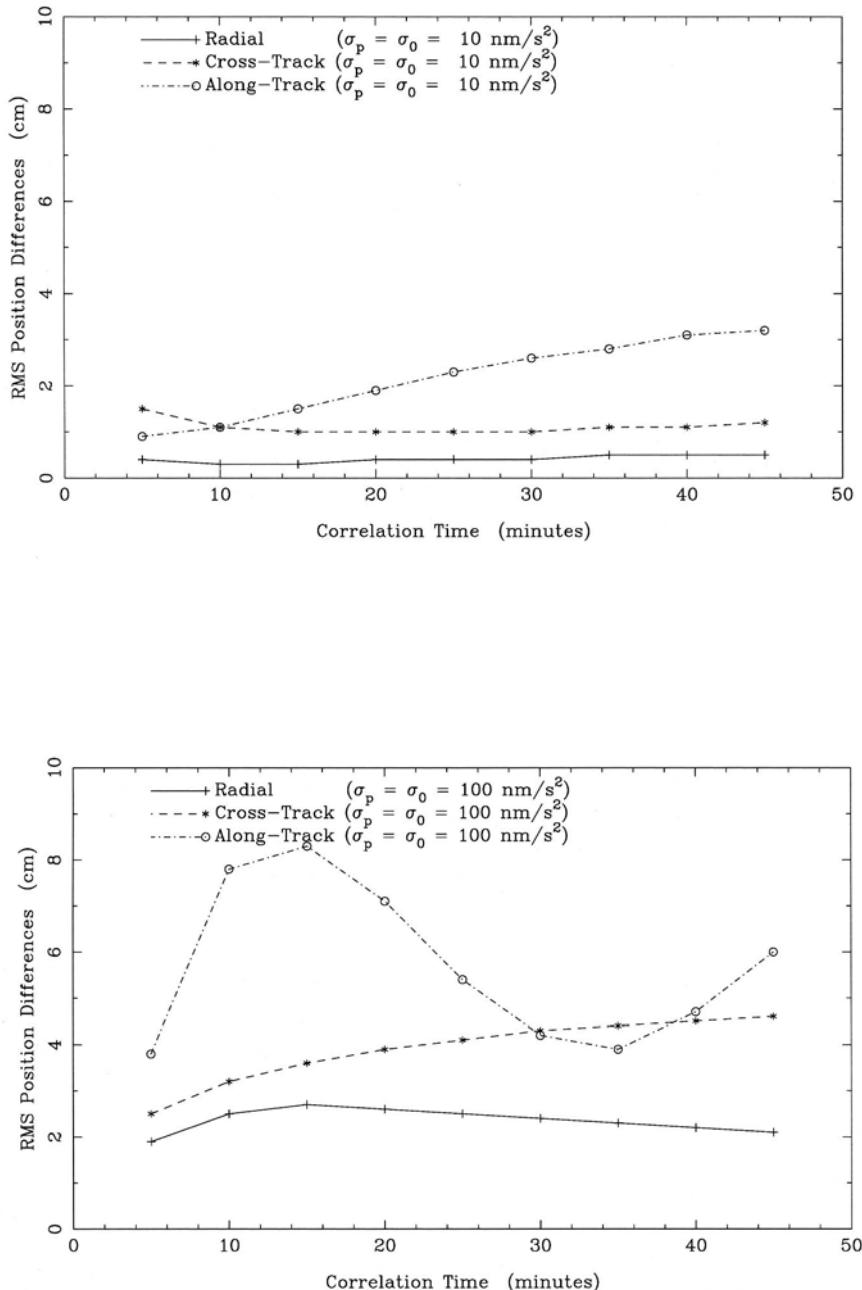


Figure 2. Differences with respect to “best” T/P reduced-dynamic orbit caused by different values for the correlation time and acceleration standard deviation (Gauss-Markov, Kalman filter). Results were obtained with JPL GIPSY/OASIS software using the JGM1 gravity field model.

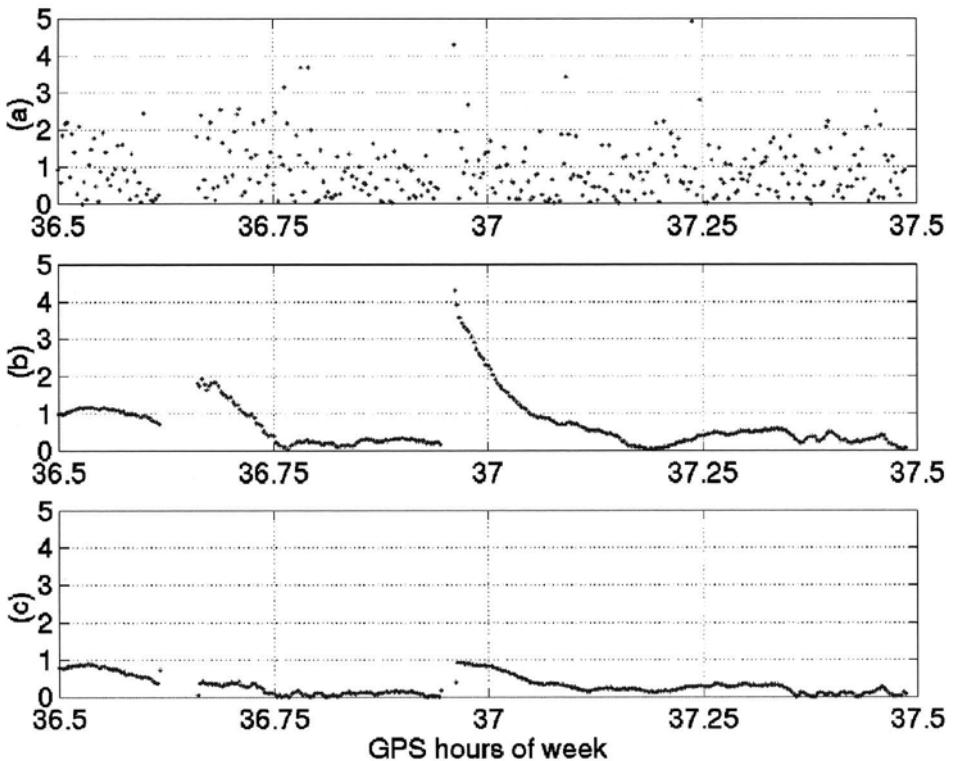


Figure 3. Comparison of T/P orbit obtained by a sequential filter kinematic technique and high-precision JPL POE for 13 November 2000. The rms of 3-dimensional orbit errors is equal to 154 cm when using pseudo-range observations (top), 43 cm in the forward filter step using phase observations (middle) and 33 cm after backward smoothing (bottom). Results were provided by S. Bisnath, New Brunswick (values along vertical axis in m).

accelerometer biases and scale factors was not addressed. Recent experiences with the CHAMP accelerometer data show that this issue needs further attention. More details of the GOCE simulations can be found in (Visser and van den IJssel, 2000).

3.2. CURRENT MISSIONS

High-precision orbit determination is now the focus of activities related to, e.g., the Jason-1 satellite (GPS, DORIS, SLR), ENVISAT (DORIS, SLR) and CHAMP (GPS, SLR, accelerometer). CHAMP obviously poses the biggest challenge, because it flies at very low altitudes, 450 km and below. It has to be noted that precise orbit determination is also a challenge for ENVISAT despite the fact that it flies at a much higher altitude of 800 km. This is because the modeling of non-gravitational forces is very complicated due to its massive dimensions. The focus of this section is on CHAMP POD activities, because kinematic POD for ENVISAT

TABLE I
DEOS CHAMP GPS-based reduced-dynamic orbit determination

DOY 2001	6-hour overlaps (cm)				SLR rms (cm)	
	Radial	Along-track	Cross-track	Norm	DOY 2001	SLR rms
140–141	1.46	2.03	2.03	3.22	140	5.26
141–142	2.04	3.32	1.88	4.33	141	5.01
142–143	1.74	3.00	2.52	4.29	142	4.88
143–144	1.25	2.39	1.49	3.08	143	3.17
144–145	1.85	2.89	1.69	3.83	144	6.33
145–146	0.99	1.68	2.16	2.91	145	3.61
146–147	1.56	2.65	2.90	4.23	146	2.52
147–148	1.75	1.41	1.61	2.77	147	4.96
148–149	0.91	1.76	1.38	2.42	148	5.55
149–150	1.16	1.85	1.66	2.74	149	3.49

Comparison with external orbits				
Institute	Radial	Along-track	Cross track	Norm
TUM	8.81	9.53	8.71	15.63
UT/CSR	2.67	6.01	4.21	7.81

is impossible, since it is not equipped with a GPS receiver. These activities were carried out predominantly in the framework of the IGS/LEO POD Pilot Project (IGSLEO PP) (<http://nng.esoc.esa.de/gps/igsleo.html>, accessed May 2002).

Reduced-dynamic orbits have been computed by several institutes, including the Delft Institute for Earth-Oriented Space Research (DEOS). The DEOS reduced-dynamic POD approach is based on ionospheric-free triple-difference GPS observations with a ground network of 50 GPS stations and the estimation of radial, along-track and cross-track accelerations in 30-min intervals, *cf.* (Rim et al., 2001). The a priori dynamic model included the TEG-4 gravity field model (Tapley et al., 2001). The rms of fit of the observations is about 0.3 mm/s (30 second time interval). The IGSLEO Pilot Project test period covers DOY 140–150, 2001. The orbit arc length was selected to be equal to 30 hours resulting in 6-hour overlaps between consecutive orbit solutions. The rms of 3-dimensional overlap orbit differences (status 8 March 2002) was found to be at the 3 cm level and the comparison of the GPS-derived CHAMP orbits with SLR observations (which were not used in the POD) results in rms of fit values of around 5 cm, including all laser stations which tracked CHAMP in the relevant time frame (Table I).

Comparisons were made between reduced-dynamic orbits computed by the Technical University of Munich (TUM) and the Center for Space Research of the

University of Texas (UT/CSR). These orbits were all computed or provided in the same time frame (January 2001 – March 2002) and are of good quality enabling a “fair” comparison (CHAMP orbit accuracy improved rapidly in the first months of 2002). The 3-dimensional orbit differences are at the 15 cm (TUM) and 8 cm level (CSR) indicating that it is fair to assume that orbit solutions are converging to within the decimeter level. Moreover, TUM is currently (March 2002) computing kinematic orbit solutions for CHAMP that are in close agreement with its reduced-dynamic POD solutions (M. Rothacher and D. Svehla, personal communication). One of the questions that needs to be addressed is whether orbit accuracy and consistency can be improved by including the CHAMP accelerometer observations in the POD. This issue currently attracts much attention, but is yet unresolved. Issues like accelerometer bias and scale factor estimation are under review and algorithms are updated continuously.

4. Conclusions and Outlook

Reduced-dynamic and kinematic precise orbit determination of LEO satellites have evolved into mature techniques. The feasibility and capability of these techniques have been clearly demonstrated by missions in the past and by the existing CHAMP satellite. GPS-based LEO orbit accuracies are rapidly approaching the cm level.

In order to further improve orbit accuracy, clear challenges can be identified in the field of carrier phase ambiguity resolution. Moreover, more attention needs to be paid to optimal parameter estimation schemes (such as keeping GPS orbits fixed or estimating them simultaneously with the LEO orbit). Other issues that deserve more attention are the inclusion of attitude and accelerometer observations in the POD. In addition, a satellite like GOCE will be equipped with a Drag Free Control (DFC) system. DFC information might also be included and help improving orbit accuracy. Concerning GRACE, it is interesting to assess the possible impact of using low-low SST in the POD and to apply space-borne GPS differencing schemes.

Acknowledgements

The GeoForschungsZentrum, Potsdam, Germany, kindly provided the high-quality GPS observations. The Center for Space Research, Austin, Texas, is acknowledged for providing the TEG-4 gravity field model. Thanks go to Henno Boomkamp (ESOC) for coordinating the CHAMP information flow in the framework of the IGSLEO Pilot Project.

References

- Bisnath, S.B., and R.B. Langley (2001), High-precision Platform Positioning with a Single GPS Receiver, in *ION GPS 2001, 11–14 September 2001, Salt Lake City, Utah, U.S.A.*, 9 pp.
- Bock, H., G. Beutler, and U. Hugentobler (2001), Kinematic Orbit Determination for Low Earth Orbiters (LEOs), *IAG 2001 Scientific Assembly, 2–7 September 2001, Budapest, Hungary, CD-rom, 6 pp.*
- Byun, S.H., and B.E. Schutz (2001), Improving satellite orbit solution using double-differenced GPS carrier phase in kinematic mode, *J. Geod.* **75**(9/10), 533–543.
- ESA (1999), Gravity Field and Steady-State Ocean Circulation Mission, *Reports for Mission Selection, The Four Candidate Earth Explorer Core Missions, SP-1233(1)*, European Space Agency, July 1999.
- NRC (1997), *Satellite Gravity and the Geosphere: Contributions to the Study of the Solid Earth and Its Fluid Envelopes*, Committee on Earth Gravity from Space, National Research Council, National Academy Press, ISBN 0-309-05792-2.
- Reigber, Ch., P. Schwintzer, and H. Lühr (1999), The CHAMP geopotential mission, in *Bollettino di Geofisica Teorica ed Applicata, Vol. 40, No. 3–4, Sep.–Dec. 1999, Proceedings of the 2nd Joint Meeting of the International Gravity and the International Geoid Commission, Trieste 7–12 Sept. 1998, ISSN 0006-6729*, edited by I. Marson and H. Sünkel, pp. 285–289.
- Rim, H., Z. Kang, P. Nagel, S. Yoon, S. Bettadpur, B. Schutz, and B. Tapley (2001), CHAMP Precision Orbit Determination, *AAS/AIAA 2001 AAS-01-334*.
- Smith, A. J. E., E. T. Hesper, D. C. Kuijper, G. J. Mets, P.N.A.M. Visser, B. A. C. Ambrosius, and K. F. Wakker (1994), TOPEX/POSEIDON Data Analysis Study, *Final Report, ESOC contract 10261/92/D/CS*, 163 pp., Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands, September 1994.
- Svehla, D., and M. Rothacher (2001), Kinematic Orbit Determination of LEOs Based on Zero or Double-difference Algorithms Using Simulated and Real SST Data, *IAG 2001 Scientific Assembly, 2–7 September 2001, Budapest, Hungary, CD-rom, 7 pp.*
- Tapley, B. D., J.C. Ries, G.W. Davis, R.J. Eanes, B.E. Schutz, C.K. Shum, M.M. Watkins, J.A. Marshall, R.S. Nerem, B.H. Putney, S.M. Klosko, S.B. Luthcke, D. Pavlis, R.G. Williamson, and N.P. Zelensky (1994), Precision orbit determination for TOPEX/POSEIDON, *J. Geophys. Res.* **99**(C12), 24,383–24,404.
- Tapley, B.D., et al. (1994), The JGM-3 gravity model, *Ann. Geophys.* **12**, suppl. 1, C192.
- Tapley, B.D., S. Bettadpur, D.P. Chambers, M.K. Cheng, B.C. Gunter, Z. Kang, J. Kim, P. Nagel, J.C. Ries, H.J. Rim, P.J. Roessel, and I. Roundhill (2001), Gravity field determination from CHAMP using GPS tracking and accelerometer data: initial results, in *AGU Fall'01 Meeting, Abstract G51A-0236*.
- Visser, P.N.A.M., and J. van den IJssel (2000), GPS-based precise orbit determination of the very low Earth orbiting gravity mission GOCE, *J. Geod.* **74**(7/8), 590–602.
- Wu, S.C., T.P. Yunk, and C.L. Thornton (1990), Reduced-Dynamic Technique for Precise Orbit Determination of Low Earth Satellites, *J. Guidance* **14**(1), 24–30.

Address for Offprints: Dr. ir. P.N.A.M. Visser, Delft Institute for Earth-Oriented Space Research, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands, Phone: +31 15 2782595, Fax: +31 15 2785322, email: Pieter.Visser@lr.tudelft.nl

SPACE-WISE, TIME-WISE, TORUS AND ROSBOROUGH REPRESENTATIONS IN GRAVITY FIELD MODELLING

NICO SNEEUW (sneeuw@ucalgary.ca)

Department of Geomatics Engineering, University of Calgary

Received: 3 June 2002; Accepted in final form: 26 September 2002

Abstract. The *decade of the geopotentials* started July 2000 with the launch of the German high-low SST mission CHAMP. Together with the joint NASA-DLR low-low SST mission GRACE and the ESA gradiometry mission GOCE an unprecedented wealth of geopotential data becomes available over the next few years.

Due to the sheer number of unknown gravity field parameters (up to 100 000) and of observations (millions), especially the latter two missions are highly demanding in terms of computational requirements. In this paper several modelling strategies are presented that are based on a semi-analytical approach. In this approach the set of normal equations becomes block-diagonal with maximum block-sizes smaller than the spherical harmonic degree of resolution. The block-diagonality leads to a rapid and powerful gravity field analysis tool.

Beyond the more-or-less conventional space-wise and time-wise formulations, the torus approach and Rosborough's representation are discussed. A trade-off between pros and cons of each of the modelling strategies will be given.

Keywords: Gravity field representations, space-wise, time-wise, Kaula, torus approach, Rosborough, CHAMP, GRACE, GOCE

1. Introduction

Gravity field recovery up to a certain maximum spherical harmonic degree L involves L^2 unknown spherical harmonic coefficients C_{lm} and S_{lm} (or complex harmonics K_{lm}). The corresponding number of normal matrix entries is L^4 . For CHAMP, GRACE and GOCE the maximum degree of resolution will—roughly—be 75, 150 and 300, respectively. Thus the number of unknowns for GOCE will be close to 100 000, for example. At the same time, the number of observations made during an entire mission will be millions. Both CHAMP and GRACE will fly for 5 years.

Clearly, gravity field recovery from these dedicated gravity missions is a demanding task. Many authors have contributed to strategies for solving such huge normal equation systems, e.g., (Schuh, 1996; Visser et al., 2001; Ditmar & Klees, 2002).

The normal matrices are generally not sparse. But after rearranging the parameters according to spherical harmonic order m , the systems turn out to become block-diagonal dominant. This has led to the so-called semi-analytical approach that yields true block-diagonal matrices, e.g. (Schrama, 1991; Koop, 1993) or



(Sneeuw, 2000), resulting in a highly efficient inversion of the normal system. The cost, though, is reduced accuracy that may be regained by iteration (Klees et al., 2000).

A key element of the semi-analytical approach is the constant radius approximation. This means that data are projected onto an ideal geometric configuration with constant radius. This can be an orbit (1D), a sphere (2D) or even a torus (2D). For each of these configurations, the *representation* of the gravity field functionals is different.

Classically, a distinction has been made between *space-wise* and *time-wise* approaches, (Rummel et al., 1993). The former models the observable as a function of spatial coordinates, leading to spherical harmonic analysis of data on a sphere. The latter treats the data as a time-series along the orbit, leading to Fourier analysis, with an implicit or explicit link between the Fourier and the spherical harmonic coefficients. More recently the time-wise approach was converted by Sneeuw (2001) into a *torus-based* method which has characteristics of both the time- and space-wise approach.

As a fourth basic representation the *Rosborough* approach should be mentioned. It separates the orbital data into ascending and descending orbit segments, cf. (Rosborough, 1986). It is a space-wise approach, though with a new set of base functions that depend on the type of the gravity functional and on the orbit. It is therefore similar to spherical harmonic analysis but now with data on two spheres.

In this paper the various representations are reviewed in a unifying way. Time-wise, space-wise, torus and Rosborough approaches are connected in a natural way. The Rosborough approach will be emphasized in particular. In literature it has been treated in a less pronounced way, so far, mostly because of algorithmic aspects that rendered the Rosborough approach cumbersome. A fast algorithm will be presented.

2. Gravity Field Representations

The starting point of all representations to follow is a spherical harmonic (SH) series of the gravitational potential:

$$V(r, \phi, \lambda) = \frac{GM}{R} \sum_{l=0}^L \sum_{m=-l}^l \left(\frac{R}{r}\right)^{l+1} K_{lm} P_{lm}(\sin \phi) e^{im\lambda}. \quad (1)$$

Although not relevant, it is assumed that the SH coefficients and functions are fully normalized. What is relevant, though, is the use of complex valued quantities. Not only does this result in a far more compact formulation, it also significantly alleviates the transition between the different representations. For definitions refer to (Sneeuw, 2000). Note that the SH series runs to the maximum degree L .

2.1. REPRESENTATION ON A SPHERE: SPACE-WISE

Any isotropic functional f of the geopotential (1) can be expressed as:

$$f(\phi, \lambda, r) = \sum_{l=0}^{\infty} \sum_{m=-l}^l h_l^f(r) K_{lm} P_{lm}(\sin \phi) e^{im\lambda}, \quad (2a)$$

or, after swapping the l - and m -summation in a two-stage formulation:

$$f(\phi, \lambda, r) = \sum_{m=-L}^L A_m^f(\phi, r) e^{im\lambda}, \quad (2b)$$

$$A_m^f(\phi, r) = \sum_{l=|m|}^L h_l^f(r) P_{lm}(\sin \phi) K_{lm}, \quad (2c)$$

in which h_l^f can be considered as a spectral transfer coefficient containing a dimensioning factor, upward continuation and the specific transfer (or eigenvalue) of the isotropic operator $V \mapsto f$.

The basic data processing scheme, corresponding to this two-stage formulation, consists of the following steps:

- Introduce a nominal sphere (constant radius approximation). The dependence in (2b) and (2c) on the radial variable r vanishes.
- Project onto the nominal sphere, for instance at mean satellite altitude and interpolate onto spherical grid .
- Perform a 1D Fourier analysis along parallels, leading to *latitude-lumped coefficients* $A_m^f(\phi)$.
- Invert the linear equation system (2c) by a block-wise least squares adjustment.
- Iterate to mitigate projection and interpolation errors.

Block-diagonality is essentially achieved by the two-stage approach in conjunction with a regular spherical grid at constant radius. The linear system (2c) can be inverted for each order m individually.

2.2. REPRESENTATION ALONG AN ORBIT: TIME-WISE

The gravitational potential (1) is transformed now into the orbital coordinates r (radius), $u = \omega + \nu$ (argument of latitude), I (inclination) and $\Lambda = \Omega - \text{GAST}$ (longitude of ascending node), cf. Figure 1:

$$V(r, u, I, \Lambda) = \frac{GM}{R} \sum_{l=0}^L \sum_{m=-l}^l \sum_{k=-l}^l \left(\frac{R}{r} \right)^{l+1} K_{lm} F_{lmk}(I) e^{i(ku+m\Lambda)}. \quad (3)$$

This expression is equivalent to Kaula's representation (1966), with the distinction that complex quantities have been used, implying modified inclination functions

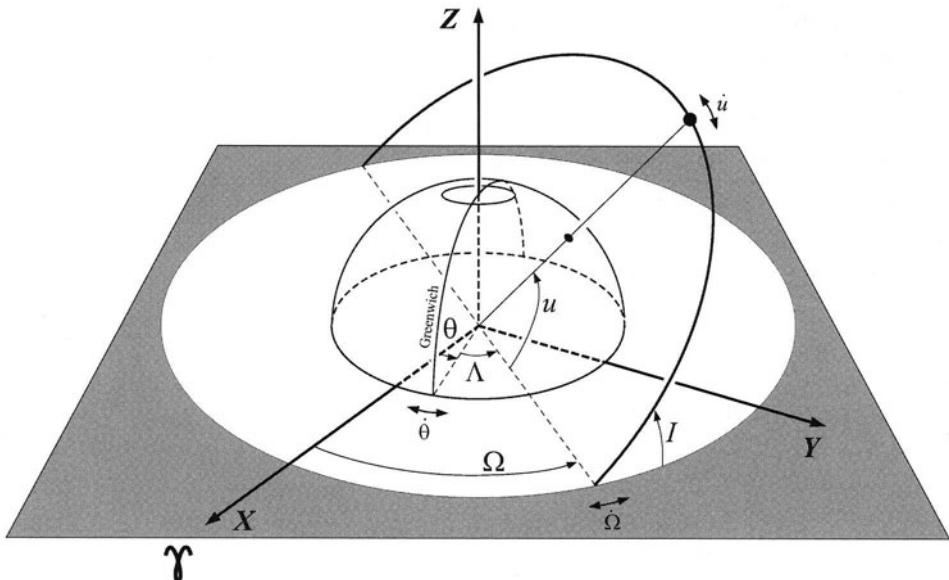


Figure 1. Orbit configuration and graphical explanation of orbital variables.

$F_{lmk}(I)$. Moreover, eqn. (3) is not expanded in eccentricity. Thus only the four aforementioned variables are used.

The mapping $V \mapsto f$ is represented in the spectral domain by a specific transfer h_{lmk}^f . In this case f is not limited to isotropic functionals. Combined with a dimensioning factor, an upward continuation term and an inclination function, we obtain the *transfer coefficient* H_{lmk}^f . Every functional is characterized spectrally by its own transfer coefficient. The corresponding series expression of the functional f can be written in a two-stage formulation too:

$$f(u, \Lambda, r, I) = \sum_m \sum_k A_{mk}^f(r, I) e^{i(ku + m\Lambda)}, \quad (4a)$$

$$A_{mk}^f(r, I) = \sum_l H_{lmk}^f(r, I) K_{lm} . \quad (4b)$$

The coefficients A_{mk}^f are known as *lumped coefficients*. Equations (4a) and (4b) lead to an efficient algorithm if we use the concept of a *nominal orbit* of constant inclination I and constant radius r . If we neglect these variables, eqn. (4b) becomes a linear system with constant coefficients. Note again that we can define such a linear system for each order m individually, leading to a block-diagonal normal matrix.

The basic data processing (at least in the frequency domain) scheme consists of the following steps:

- Introduce a nominal orbit (constant radius approximation and constant inclination).

- Project the data onto the nominal orbit using prior gravity field information.
- Perform a 1D Fourier over the repeat period, leading to Fourier coefficients A_n^f .
- Map the 1D Fourier spectrum onto the *lumped coefficients* A_{mk}^f .
- Invert (4b) by block-wise least squares.
- Iterate the process to mitigate projection errors.

This approach, in which the lumped coefficients were obtained explicitly by the Fourier transformation, is known as the *time-wise approach in the frequency domain*. Alternatively, one can follow the *time-wise approach in the time domain*. Without the two-stage formulation, we would have a linear relationship between $f(t)$ and K_{lm} directly. In setting up the corresponding normal equations, a Fourier transform is performed implicitly, leading to the same normal system, cf. (Koop, 1993). The requirements for both time-wise approaches are a repeat orbit and a regularly sampled data set.

2.3. REPRESENTATION ON A TORUS

Equation (4a) reveals the character of the lumped coefficients A_{mk}^f : they are the 2D Fourier coefficients of the functional $f(u, \Lambda)$, cf. (Sneeuw, 2001). Since both, u and Λ , are defined on the whole circle, f is a periodic function on a torus, which is the topological product of two circles. Neglecting periodic orbit perturbations ground-tracks are straight lines in the (u, Λ) -domain. Figure 2 illustrates these ideas using the radial gravity gradients—one of the GOCE observables.

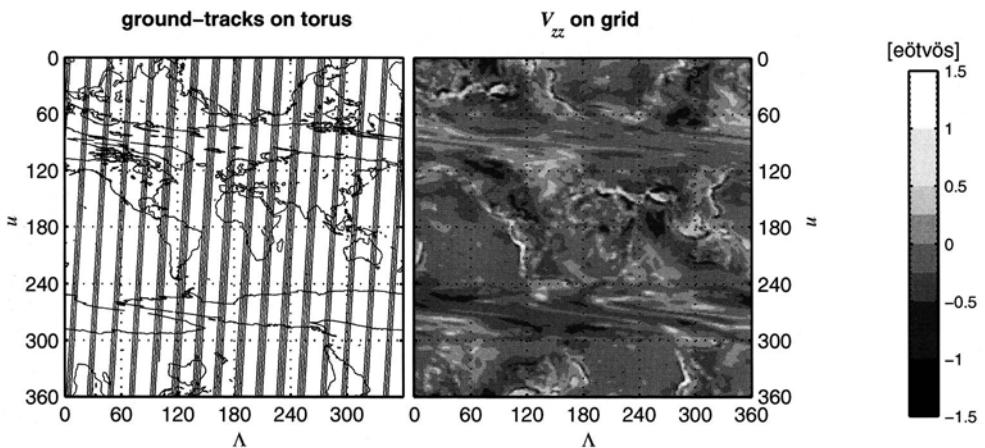


Figure 2. Some orbit revolutions (left) and radial gravity gradient (right) on torus. The Earth is represented twice, apart from polar gaps, since it will be covered by ascending and descending tracks which are separated naturally in the torus approach.

First, the data might have to be pre-processed to guarantee periodicity. The basic data processing scheme then consists of the following steps:

- Project and interpolate onto torus of constant radius $\Rightarrow f(u, \Lambda)$.
- Perform a 2D Fourier analysis \Rightarrow lumped coefficients A_{mk}^f .
- Invert eqns. (4b) by block-wise least squares adjustment.
- Iterate to mitigate projection and interpolation errors.

The requirement of a repeat orbit has been dropped and data gaps pose no problem anymore. This allows for far more flexible sampling situations. The cost, however, is the need to interpolate on the torus and, hence, interpolation errors.

2.4. REPRESENTATION ON TWO SPHERES: ROSBOROUGH

The starting point of our considerations was a SH series, i.e., a representation in spherical coordinates. For the time-wise and torus approach it was transformed into representations in orbital elements. Rosborough (1986) performed the reverse transformation. He managed to express the along-orbit observable, in particular orbit perturbations, in spherical coordinates again. However, since the observable is non-isotropic in general, the result will not be a simple SH series. New base functions are required that depend on the type of observable and on the nominal orbit. Moreover one has to distinguish between ascending and descending arcs.

The key element in Rosborough's transformation is to express $e^{i(ku+m\Lambda)}$ in (4a) in terms of spherical coordinates ϕ and λ . At this point the advantage of using complex quantities becomes very clear. In (Rosborough, 1986) the sines and cosines are developed into binomial series, leading to a cumbersome derivation and a slow algorithm. There is no need for that in complex notation.

First, we consider the longitude part. In the Earth-fixed frame, the longitude of the satellite is λ and the longitude of the ascending node is Λ . In the inertial frame, the right ascension of the satellite is α and that of the ascending node is Ω . Thus we arrive at:

$$\lambda - \Lambda = \alpha - \Omega \implies e^{i(ku+m\Lambda)} = e^{i(ku-m(\alpha-\Omega))} e^{im\lambda}. \quad (5)$$

In the next step we need to transform the variables u and $(\alpha - \Omega)$ into expressions in latitude ϕ and inclination I . From basic spherical trigonometry it is:

$$\cos u = \cos \phi \cos(\alpha - \Omega), \quad (6a)$$

$$\sin u = \cos \phi \sin(\alpha - \Omega) / \cos I, \quad (6b)$$

$$\sin u = \sin \phi / \sin I. \quad (6c)$$

From the latter, one obtains:

$$\cos u = \pm \frac{1}{\sin I} \sqrt{\sin^2 I - \sin^2 \phi}, \quad (6d)$$

where the + sign is valid for ascending tracks ($u \in [-\frac{1}{2}\pi, \frac{1}{2}\pi]$) and the - sign for descending tracks ($u \in [\frac{1}{2}\pi, \frac{3}{2}\pi]$). Note that (6d) implies the condition $\sin^2 \phi <$

$\sin^2 I$, or $|\phi| < \frac{1}{2}\pi - |\frac{1}{2}\pi - I|$. Combining these equations, and making use of $e^{ix} = \cos x + i \sin x$, we end up with:

$$\begin{aligned}\Phi_{mk}^\pm(\phi, I) &= e^{i(ku-m(\alpha-\Omega))} \\ &= \frac{1}{\cos^m \phi \sin^{m+k} I} \left(\pm \sqrt{\sin^2 I - \sin^2 \phi} + i \sin \phi \right)^k \\ &\quad \cdot \left(\pm \sqrt{\sin^2 I - \sin^2 \phi} - i \sin \phi \cos I \right)^m.\end{aligned}\quad (7)$$

The function $\Phi_{mk}^\pm(\phi, I)$ is just a shorthand expression for the right hand side of eqn. (7). The superscript \pm denotes that it is different for ascending and descending tracks. Despite the powers k and m eqn. (7) gives rise to a numerically stable and fast algorithm for computing $\Phi_{mk}^\pm(\phi, I)$.

These functions are used to define new base functions $Q_{lm}^\pm(\phi, I)$ that replace the Legendre functions of (1). Also these functions will depend on inclination and will be different for ascending and descending tracks:

$$f^\pm(\phi, \lambda) = \sum_l \sum_m K_{lm} Q_{lm}^\pm(\phi, I) e^{im\lambda}, \quad (8a)$$

$$\text{with: } Q_{lm}^\pm(\phi, I) = \sum_k H_{lmk}^f \Phi_{mk}^\pm(\phi, I). \quad (8b)$$

Equation (8a) is very similar to a SH series, although the $Q_{lm}^\pm(\phi, I)$ are in general not orthogonal. The data processing strategy in the Rosborough approach is therefore very similar to that of the space-wise approach. The main difference is that we have two spheres now: one for the data on ascending tracks and one for the data on descending tracks. In practice this is often combined into spatially *mean* and *variable* contributions: $f^m = \frac{1}{2}(f^+ + f^-)$ and $f^v = \frac{1}{2}(f^+ - f^-)$. The spatially mean contribution is misleadingly known as *geographically correlated* as well. The data processing scheme becomes:

- Separate data from ascending and descending tracks.
- Interpolate onto two spheres $\Rightarrow f^+(\phi, \lambda)$ and $f^-(\phi, \lambda)$.
- Create *mean* and *variable* spherical functions $\Rightarrow f^m(\phi, \lambda)$ and $f^v(\phi, \lambda)$.
- Invert by block-wise least squares $\Rightarrow K_{lm}$.
- Iterate.

Figure 3 illustrates these spatially mean and variable fields, using the GOCE gravity gradient observables. To be precise, the gravity gradients of the *disturbing* potential are visualized. The three fields on the left are spatially mean contributions. At the right, one can see the corresponding spatially variable parts. Observe the different grayscales left and right. Since the GOCE orbit is close to polar, the along-track direction on an ascending track is close to the negative along-track direction on the descending track. The same holds for cross-track. This explains why the spatially variable fields at the right are so much smaller in magnitude. The maximum orientation variation—and therefore f^v —will occur at the high latitudes.

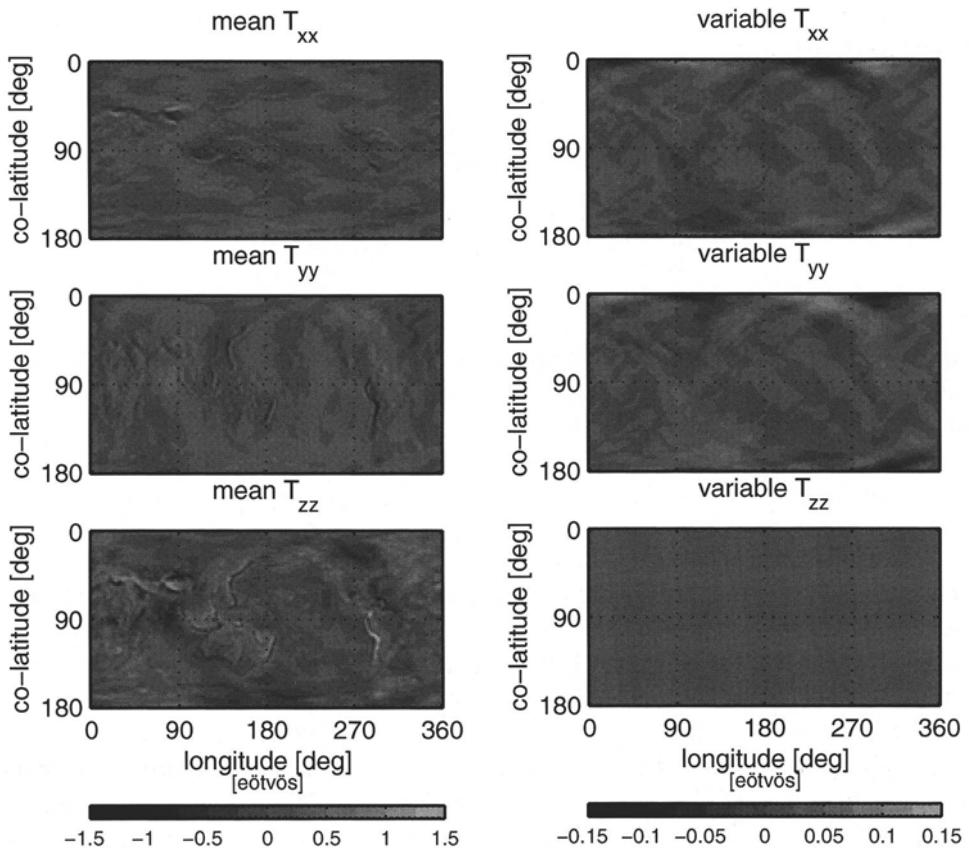


Figure 3. GOCE gravity gradients: spatially mean (left) and variable (right). Note the different scales. Coordinates are along-track (x), cross-track (y) and radial (z).

The radial gradient T_{zz} is isotropic. It does not depend on direction (ascending or descending). Therefore the spatially variable component must be zero.

3. Trade-off and Conclusions

In this paper all gravity field representations have been derived from the spherical harmonic series (1). Therefore, they are all equivalent. Only the space-wise approach itself cannot deal with non-isotropic functionals easily.

Each representation leads to a data processing method, all of which are feasible for gravity field satellite missions. The method of choice depends on practical considerations, like the orbit configuration (repeat orbit or not), data stream continuity (data gaps, thruster events, orbit manoeuvres). It depends, moreover, on the noise characteristics (given in spectral, temporal or spatial domain?), on the observable being isotropic or not, on its interpolation properties, and so on. The strengths

and weaknesses of the different representations and corresponding methods are summarized in table I.

TABLE I
Strengths (\sharp) and weaknesses (\flat) of the various approaches
and representations

	space	time	freq.	torus	Rosb.
data type	\flat	\sharp	\sharp	\sharp	\sharp
repeat orbit	\sharp	\flat	\flat	\sharp	\sharp
interpolation	\flat	\sharp	\sharp	\flat	\flat
data gaps	\sharp	\flat	\flat	\sharp	\sharp
spectral stoch.	\flat	\flat/\sharp	\sharp	\sharp	\flat/\sharp

For example, the GOCE gradients—except for V_{zz} —are non-isotropic. The noise characteristics of the gradiometer are given in terms of error power spectral densities. If GOCE would fly a repeat orbit, the time-wise approach in the frequency domain would be a good choice. If not, it would make sense to resort to the torus approach. That would also be useful in the case of data gaps.

It was shown that the use of complex-valued quantities is not only beneficial to the compactness of the formulae derived, but also to revealing the character of lumped coefficients (2D Fourier coefficients of the functional on a torus) and particularly for deriving the Rosborough transformation and simplifying the corresponding algorithm.

Acknowledgements

I would like to thank Björn Frommknecht, Technical University of Munich, for his testing and programming help on the Rosborough approach. I am very grateful to Georges Balmino for discussing the improved algorithm for the Rosborough approach. In an unpublished memorandum Balmino (1996) developed the Rosborough transformation along the same lines, i.e., in complex notation and without binomial series.

References

- Balmino, G.: 1996, 'A Note on Rosborough transformation', unpublished memorandum.
 Ditmar P. R. Klees.: 2002 'A Method to Compute the Earth's Gravity Field from SGG/SST Data to be Acquired by the GOCE Satellite', DUP Science, Delft, The Netherlands

- Kaula, W.M.: 1966, *Theory of satellite geodesy*, Blaisdell Publishing Co.
- Koop, R.: 1993, Global Gravity Field Modelling Using Satellite Gravity Gradiometry, Publications on Geodesy, New Series **38**, Netherlands Geodetic Commission
- Klees, R. R., Koop, P., Visser, and J., van den IJssel.: 2000, Fast Gravity Field Recovery From GOCE Gravity Gradient Observations, *J. Geodesy* **74**:561–571
- Rosborough, G.W.: 1986, Satellite Orbit Perturbations due to the Geopotential, **CSR-86-1**
- Rummel, R. M., van Gelderen, R. Koop, E. Schrama, F. Sansò, M. Brovelli, F. Migliaccio, F. Sacerdote.: 1993, Spherical harmonic analysis of satellite gradiometry, Publications on Geodesy, New Series **39**, Netherlands Geodetic Commission
- Schrama, E.J.O.: 1991, Gravity Field Error Analysis: Applications of Global Positioning System Receivers and Gradiometers on Low Orbiting Platforms, *J. Geophys. Res.* **96(B12)**:20 041–20 0051
- Schuh, W.D.: 1996, Tailored Numerical Solution Strategies for the Global Determination of the Earth's Gravity Field, Mitteilungen der Universität Graz, Folge **81**
- Sneeuw, N.: 2000, A semi-analytical approach to gravity field analysis from satellite observations, Deutsche Geodätische Kommission, Reihe C, **Nr. 527**
- Sneeuw, N.: 2001, Satellite Geodesy on the Torus — Block-Diagonality from a Semi-Analytical Approach, in: M Sideris (ed.) *Gravity, Geoid, and Geodynamics 2000*, IAG symposium **123**, pp. 137–142, Springer Verlag
- Visser, P.N.A.M., J. van den IJssel, R. Koop, and R. Klees.: 2001, Exploring gravity field determination from orbit perturbations of the European Gravity Mission GOCE, *J. Geodesy* **75**:89–98

Address for Offprints:

Nico Sneeuw

Department of Geomatics Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada, T2N 1N4

GRAVITY FIELD RECOVERY FROM GRACE: UNIQUE ASPECTS OF THE HIGH PRECISION INTER-SATELLITE DATA AND ANALYSIS METHODS

G. BALMINO

Centre National d'Etudes Spatiales, GRGS, Toulouse, France (georges.balmino@cnes.fr)

Received: 17 July 2002; Accepted in final form: 26 September 2002

Abstract. The very high accuracy of the Doppler and range measurements between the two low-flying and co-orbiting spacecraft of the GRACE mission, which will be at the $\mu\text{m/sec}$ and $\sim 10 \mu\text{m}$ levels respectively, requires that special procedures be applied in the processing of these data. Parts of the existing orbit determination and gravity field parameters retrieval methods and software must be modified in order to fully benefit from the capabilities of this mission. This is being done in the following areas: (i) numerical integration of the equations of motion (summed form, accuracy of the predictor-corrector loop, Encke's formulation); (ii) special inter-satellite dynamical parameterization for very short arcs; (iii) accurate solution of large least-squares problems (normal equations vs. orthogonal decomposition of observation equations); (iv) handling the observation equations with high accuracy. Theoretical concepts and first tests of some of the newly implemented algorithms are presented.

Keywords: low-low satellite-to-satellite tracking, global gravity modelling, numerical integration, large inverse problem

1. Introduction

The GRACE mission started on March 17, 2002 with the launch of two co-orbiting satellites on a quasi-polar orbit ($\sim 89^\circ$ inclination) of low mean altitude (~ 480 km, to decay naturally under the upper atmosphere drag). The main purpose of the mission is, over a three to five year life time, to determine the time variations of the Earth's gravity field every month and with a spatial resolution of about 300 km, also to contribute to the modelling of the mean field with unprecedented accuracy down to about 150 km resolution. It is the second dedicated gravity mapping mission of this decade, after CHAMP, which has been successfully operating for about two years (at the time this paper is written), and before GOCE, which is planned to fly between 2006 and 2008. The expected performances of CHAMP and GRACE, compared to those of a recent satellite solution (GRIM5-S1) of the previous decade, are shown on Figure 1. For doing so, the two spacecraft are tracked by the GPS (Global Positioning System) satellites where pseudo-distance and millimeter-class accuracy phase data are obtained (between each GRACE spacecraft and up to eight GPS satellites), and the distance between the two co-orbiters is very accurately measured by a new K-band two-way radar system (KBR); this is exemplified by



Space Science Reviews **108:** 47–54, 2003.

© 2003 Kluwer Academic Publishers.

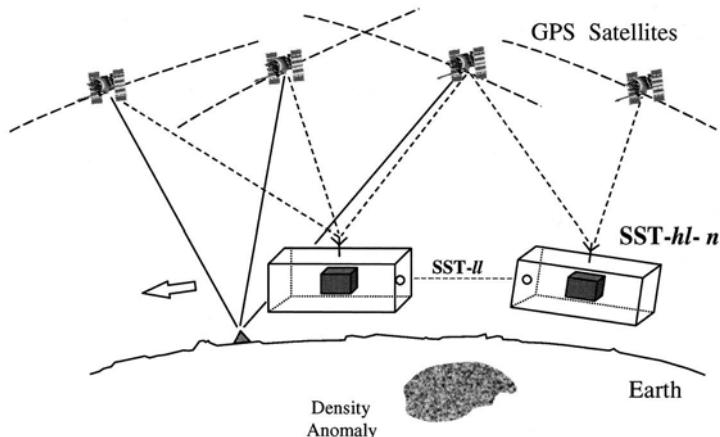


Figure 1. The GRACE mission concept: a combination of the multi-high-low (*hl-n*) and low-low (*ll*) satellite-to-satellite tracking techniques (SST).

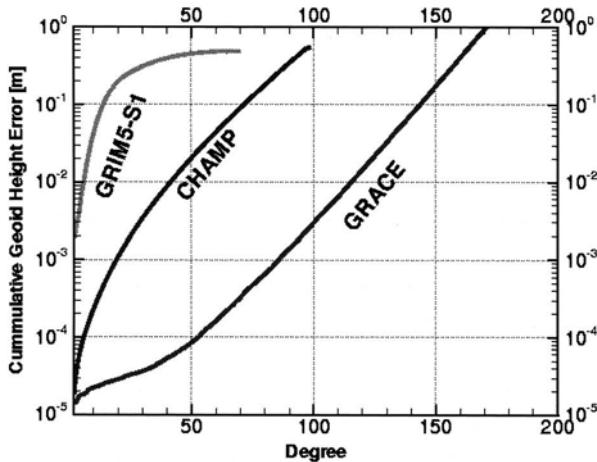


Figure 2. The expected performances of the CHAMP and GRACE missions in mapping the mean (static) gravitational potential of the Earth, in terms of geoid heights.

Figure 2. In addition, each GRACE spacecraft is equipped with a 3-axis micro-accelerometer (Super-STAR) which measures the surface accelerations: air drag, solar radiation pressure (direct and indirect — due to the Earth's albedo and infrared emission), at the level of a few $10^{-10} \text{ m s}^{-2}$. The accuracy of the KBR range measurements is about $10 \mu\text{m}$, and the relative velocity between the two orbiters can be derived at the $\mu\text{m/sec}$ level. This poses new problems of numerical metrology nature, which demand to revisit some of the methods and algorithms which are in use in the precise orbit determination and gravity field retrieval software. The critical areas in which improvements or new assessments of the obtainable accuracy are necessary have been identified; they are: (i) numerical integration of the equations of motion; (ii) special inter-satellite dynamical parameterization for

very short arcs; (iii) accurate solution of large least-squares problems; (iv) handling the observation equations with sufficient accuracy. We will describe the current works which are being done in these four areas by several teams in the world.

2. Numerical Integration of the Equations of Motion

The precision $\varepsilon(\delta r)$ of the KBR measurements is such that the relative precision of the numerical integration of the equations of motion, which reconstruct r (radius vector) and \dot{r} for each spacecraft, must be guaranteed at the 10^{-13} (10% of $\varepsilon(\delta r)/r$) or even 10^{-14} (1%) level if one computes the orbit difference $r_2 - r_1$ and related quantities from r_1 and r_2 separately. There may be ways to directly integrate the vector $r_2 - r_1$ but not to our knowledge. Apart from using 128 bit-arithmetic, which would be very costly because such computations are not hardware-doable on actual computers, some tricks have to be found to enhance the precision of the presently used integrators.

The systems to be integrated are of the form:

$$\ddot{x} = f(x, \dot{x}, t) \quad (1)$$

or

$$\ddot{y} = g(y, t) \quad (2)$$

with proper initial conditions. The systems (1) are dealt with when the equations of motion in rectangular coordinates (x) are used, while systems (2) are encountered when the planetary Lagrange-Gauss equations are considered, the y 's being the osculating orbital elements (or equivalent non-singular elements). The reader is referred to Balmino (1974) for a review of the main numerical integration methods (Cowell, Adams-Moulton, Cowell-Kulikov, etc., all from mathematicians of the past centuries) used in orbital dynamics.

First of all, the integrator stability must be guaranteed. Assuming that we work with a first order system and that a starting procedure has generated an array of $\{y_j, \dot{y}_j\}$ values ($j < N$), the running phase (for $n \geq N$) uses a predictor-corrector scheme of stepsize h which can be written as:

$$\hat{y}_n = a_0 y_n + a_1 y_{n-1} + \dots + a_k y_{n-k} + h(b_0 \dot{y}_n + b_1 \dot{y}_{n-1} + \dots + b_k \dot{y}_{n-k}) \quad (3)$$

a_0 and b_0 are zero for the predictor, and we have different coefficients a 's and b 's for the predictor and the corrector expressions. The stability of the integrator will then depend on the roots of a polynomial whose coefficients are functions of $\{a_j\}$, $\{b_j\}$, h and $\partial g/\partial y$: these roots must all be inside the unit circle.

Then we have to minimize the truncation errors which depend on the quality of the approximation of f (or g) by the prediction-correction polynomials which underly formulas such as those in equation (3). This is studied for each class of problem and usually depends on the basic (proper and exciting) frequencies of the system. A useful trick when working with systems of type (1) is to modify

the algorithm in order to exactly integrate a periodic motion; this is realized by additional terms to the predictor-corrector formulas. Apart from that one should satisfy in satellite geodesy the rule of thumb:

$$h \leq (\text{orbital period}/L_{\max})/4 \quad (4)$$

where L_{\max} is the maximum degree (and order) of the spherical harmonics representation of the gravitational potential.

Another class of errors, the round-off errors (*r.o.e.*), are function of the computer precision ε , of k (in equation 3), h and of the number of steps (n). To reduce them requires to substitute to the normal form (3) of the predictor-corrector so-called *summed* formulas which result in a much slower growing of the errors. The *r.o.e.* for the two sets of integration formulas are:

$$\begin{aligned} (\text{r.o.e.})_{\text{Normal}} &\approx 1/2n^2\varepsilon \exp[n^2h^2B_N L/z_N]/z_N \\ (\text{r.o.e.})_{\text{Summed}} &\approx n\varepsilon(1+h^2)\exp[nh^2B_S L/z_S]/z_S \end{aligned} \quad (5)$$

where $z_{N,S} = 1 - h^2L[(a_0(N, S), b_0(N, S)], B_{N,S} = [(a_n(N, S), b_n(N, S)]$, the a 's and b 's being the proper coefficients of the normal (N) and summed (S) forms, respectively, and L is a constant. The advantage of the summed form is obvious on these formulas.

For orbits with a high eccentricity (not the case of the GRACE spacecraft !) a reduction of the *r.o.e.* may be further obtained by a change of the independent variable, t (the time), into an angle σ , for which the average distribution of the satellite positions at the numerical integrator nodes is a regular mesh. The usual choice for σ is the eccentric anomaly (or the true anomaly).

Finally, and most important, is the handling of mixed precision arithmetic in the implementation of the integrator. As said above it would not be realistic today to compute everything in quadruple precision (128 bit words) — except for verifying the overall accuracy of the software by a limited number of tests. But it was found necessary, for GRACE-type problems, to do the operations of the predictor-corrector scheme in quadruple precision while keeping the computation of the right-hand side of systems (1) or (2) — that is the computation of the forces, in double precision. When working in rectangular coordinates, it is also necessary to compute separately and in quadruple precision the “central term” (*i.e.* the acceleration vector $-GM\bar{r}/r^3$) of the mean spherical Earth (of mass M , with G being the gravitational constant); other relatively large terms may also be treated similarly, such as the dynamical flattening perturbation, or all terms involved in the pseudo-correction algorithm (by which only those terms are recomputed in the predictor sequence). A similar idea is generally applied with the so-called *Encke formulation*. It consists of integrating a differential motion, with respect to a precessing reference orbit which approximates the mean motion of the true orbit (it can be obtained analytically or by a crude first numerical integration). Details can be found in Balmino and Barriot (1990). The method works best when all

TABLE I

Numerical integration precision obtained with Cowell-Kulikov integrator (order 10). One GRACE spacecraft, 2 day arc, Earth's gravity field to $L_{\max} = 120$. Comparisons made with full quadruple precision software. Tests made on a CRAY-J90 (limitation due to non IEEE arithmetic)

Type of system & algorithm	POSITION	VELOCITY
(1): rectang. elem.	10 to 100 mm	100 $\mu\text{m/sec}$
+ Summed form	1 to 10 mm	1 $\mu\text{m/sec}$
+ Pred-Corr in R*16	.1 mm	0.1 $\mu\text{m/sec}$
(2) + summed form	5 μm	0.01 $\mu\text{m/sec}$
+ Pred-Corr in R*16		

computations associated with the reference orbit are done in quadruple precision, and especially with systems of type (2).

The implementation in the GRACE case of all the above features is under way and results are very promising (see Table 1).

The tests with the Encke method have not yet been completed at the time of this paper.

3. Special Inter-satellite Dynamical Parameterization

The peculiar geometry of a pair of co-orbiting satellites at a few hundred kilometer distance from each other suggests that the inter-satellite vector behaviour may be better analysed and shown if one adopts parameters which are related to both objects instead of separate parameters for each one. The purpose is to exhibit the sensitivity of low-low satellite-to- satellite measurements (SST-ll) to some components of the two spacecraft orbits (and almost no sensitivity to others), then to retrieve the gravity model spherical harmonic coefficients from the SST-ll data only and over unusually short arcs (as may be dictated by a large number of maneuvers). The first and original works in this area have been done by a NASA-GSFC team (Rowlands *et al.*, 2001). We here follow their approach.

We introduce the middle point M of the two satellite positions S_1, S_2 and replace the classical state vector P_k of twelve elements (positions and velocities, or the two sets of osculating orbital elements) by the following one:

$$Q_j = \{r_M, \phi_M, \lambda_M \cos \phi_M, \dot{r}_M, \dot{\phi}_M, \dot{\lambda}_M \cos \phi_M, \rho, \beta, \alpha \cos \beta, \dot{\rho}, \dot{\beta}, \dot{\alpha} \cos \beta\}$$

where $r_M = OM$ (O is the origin of the coordinate system $\{\Sigma\}$ in which the equations are integrated), ϕ_M is the latitude of M in $\{\Sigma\}$, λ_M is the longitude of M in $\{\Sigma\}$, $\rho = S_1S_2$, α and β are the longitude and latitude of the S_1S_2 vector in the coordinate system centered in M and parallel to $\{\Sigma\}$. Then it suffices to transform the partial derivatives in the observation equations by writing:

$$\frac{\partial(x, \dot{x})}{\partial Q_j} = \frac{\partial(x, \dot{x})}{\partial P_k} \cdot \frac{\partial P_k}{\partial Q_j} \quad (6)$$

where the matrix $\partial P_k / \partial Q_i$ is the inverse of the jacobian matrix $\partial Q_j / \partial P_k$, which is easy to compute from the definition of the new elements.

Simulations reported by Rowlands *et al.* (2001) have been made over short arcs with GRACE type data and have shown that:

- $\dot{\rho}$ and r_M are highly correlated,
- $\dot{\rho}, \beta, \dot{\beta}$ work best at relative orbit determination and gravity field recovery.

For example, a 30 day simulation with 2878 arcs (each 15 minutes long at most) was done in two steps:

- (i) the orbit was first determined using GPS data and introducing empirical accelerations (reduced dynamics approach),
- (ii) only SST low-low data were used to determine $\dot{\rho}, \beta, \dot{\beta}$ and the gravity harmonics (up to $L_{\max} = 120$).

The results show a good agreement with the reference solution computed by a classical approach (regular orbit fit and gravity inversion from thirty-one day arcs).

Therefore it seems possible to decouple the orbit determination step from the gravity retrieval task using short arcs, resulting in a big saving of CPU in the whole process.

4. Accurate Solution of Large Least Squares Problems

One aspect of the new gravity mapping missions is the large amount of observations to be processed — yielding large number of (linearized) observation equations, and the large number of parameters which must be computed. In algebraic form, we have the following system of **m** equations and **n** unknowns (we assume **n** < **m**):

$$Ax = b \quad (7)$$

with $A = \prod^{1/2} A$, $b = \prod^{1/2} b$, \prod being the weight matrix.

This system is usually solved by least squares, using the normal equation approach. The normal system associated to (7) is:

$$A^T Ax = A^T b \quad (8)$$

and $N = A^T A$ is the normal **n** × **n** matrix. Its formation requires about **m** **n**²/2 floating operations (multiplication + addition) which is often the most expensive part of the computational process; for example, one month of GRACE intersatellite one-second data yield about **m** = 2.5 10⁶ and **n** is of the order of 30 000 (for a gravity model up to degree-order 150, plus ocean tide parameters for more than 15 components — spherical harmonics up to degree and order 50 to 80 depending on the component). Then one does the Cholesky decomposition of the normal system:

$$N = R_N^T \cdot R_N \quad (9)$$

and solves equation (8) by forward and backward substitution. The basic problem with such an approach is the loss in accuracy which degrades the precision of the solution as **m** and **n** grow.

Another method which has been shown to be as efficient (it requires about the same number of operations) but much more precise relies on orthogonal transformations. One technique uses the “QR” decomposition of the A matrix:

$$A = QR \quad (10)$$

where Q is an orthogonal matrix, product of p orthogonal $\mathbf{m} \times \mathbf{m}$ matrices Q_k (p is the rank of the system: $p < \mathbf{n}$), and R is a rectangular $\mathbf{m} \times \mathbf{n}$ matrix having only an upper triangular non zero part R^* , which turns out to be identical to R_N . As above, one then solves:

$$R^{*T} y = A^T b \quad (11)$$

and

$$R^* x = y \quad (12)$$

Alternatively, y can be computed as $y = Q^T b$.

Finally the (unscaled) covariance matrix of the solution parameters is computed as:

$$\text{Cov}(x) = R^{*-1} (R^{*-1})^T \quad (13)$$

Each Q_k matrix is the transpose of a so-called H or G matrix obtained by:

- either a Householder transformation; for instance to nullify all components, except the first one, of the first column \mathbf{m} -vector (v_i) of the A matrix, one chooses:

$$H = I_{\mathbf{m}} - 2uu^T/(uu^T) \quad (14)$$

where $u = v + \varepsilon \|v\| e_1$, $e_1 = [100 \dots 0]^T$, $\varepsilon = +1$ if $v_1 \geq 0$, $\varepsilon = -1$ if $v_1 < 0$; H is a reflection in the $(\mathbf{n}-1)$ dimension sub-space S_u orthogonal to u : $Hu = -u$ and $Hz = s$ ($s \in S_u$). This results in $Hv = -\varepsilon \|v\| e_1$. So $B = HA$ is computed, then a similar transformation is applied to the sub-matrix of B formed by deleting the first row and first column, etc.

- or a Givens transformation: this is a rotation in the $\{e_I, e_J\}$ plane by which the I .th and J .th components of a column vector (v_i) of the A matrix are transformed into 1 and 0 respectively; the G matrix is defined by:

$$G_{ij} = \delta_{ij} [1 + (c - 1)(\delta_{ii} + \delta_{jj})] + (s - 1)(\delta_{ii}\delta_{jj} - \delta_{ij}\delta_{ji}) \quad (15)$$

with $c = v_I/(v_I^2 + v_J^2)^{1/2}$, $s = v_J/(v_I^2 + v_J^2)^{1/2}$.

As normal equation systems which correspond to different batches of data can be added, it is possible to add QR-decompositions too. Several research groups involved in the GRACE mission, in the USA and in Europe, will work with either or both methods.

5. Handling the Inter-satellite Observation Equations With High Accuracy

The KBR (K-band radar) is the key science instrument on GRACE. It provides with a few micrometers accuracy the SST low-low observable which is the measurement of the dual one-way range change between both spacecraft.

The observation equations are written for the inter-satellite range (with an ambiguity for each series of uninterrupted data, to be solved for), and for the relative velocity. These equations are linearized with respect to reference quantities computed from the numerically integrated equations of motion. They involve:

- the position and velocity vectors of each satellite,
- the unit vector along the inter-satellite direction (obtained from the above),
- the partial derivatives of the instantaneous position and velocity vectors with respect to the estimated parameters: the inner ones (for each arc) and the outer ones (common to all arcs). These partials are obtained by numerically integrating the variational equations for the unknown parameters.

Due to possible round-off errors in computing differences of large quantities, it is recommended to implement the observation equations in quadruple precision for the b -vector (“observed minus computed” quantities), while generating the A matrix (the partial derivatives) in double precision. The increase of computer time can be neglected with respect to the most costly parts (formation of normal equations — or QR decomposition, evaluation of the force models in numerical integration).

6. Conclusion

The correct processing of the GRACE data has required to revise and upgrade several algorithms and modules of the orbit determination and gravity parameter retrieval software packages developed by the satellite geodesy community all over the world. This is only a first step in this domain. There is no doubt that there will be successors to the GRACE and GOCE missions, which will be even more ambitious scientifically and result in even more demanding requirements in numerical software accuracy. At some stage it may be expected that even many of the underlying physical models we are using today will have to be reconsidered. It is a non-ending and very challenging area of research by itself.

References

- Balmino, G.: 1974, ‘Numerical Methods of Orbital Dynamics’, in G.E.O. Giacaglia (ed.), *Satellite Dynamics*, Proc. COSPAR-IAU-IUTAM Symposium, Sao Paulo, Brazil, 19-21 June 1974, Springer-Verlag, Berlin, pp. 50–97.
- Balmino, G. and Barriot, J. P.: 1990, ‘Numerical Integration Techniques Revisited’, *Manuscripta Geodaetica* **15**, 1–10.
- Rowlands, D. D., Ray, R. D., Chinn, D. S. and Lemoine, F. G.: 2001, ‘Short Arc Analysis of Inter-Satellite Tracking Data in a Gravity Mapping Mission’, *Journal of Geodesy*, submitted.

GLOBAL GRAVITY FIELD RECOVERY USING SOLELY GPS TRACKING AND ACCELEROMETER DATA FROM CHAMP

C. REIGBER¹, G. BALMINO², P. SCHWINTZER¹, R. BIANCALE², A. BODE¹, J.-M. LEMOINE², R. KÖNIG¹, S. LOYER², H. NEUMAYER¹, J.-C. MARTY², F. BARTHELMESES¹, F. PEROSANZ² and S. Y. ZHU¹

¹*GeoForschungsZentrum Potsdam (GFZ), Dept. I ‘Geodesy and Remote Sensing’, Potsdam,
Germany (psch@gfz-potsdam.de)*

²*Groupe de Recherche de Géodésie Spatiale (GRGS), Toulouse, France*

Received: 6 September 2002; Accepted in final form: 10 January 2003

Abstract. A new long-wavelength global gravity field model, called EIGEN-1, has been derived in a joint German-French effort from orbit perturbations of the CHAMP satellite, exploiting CHAMP-GPS satellite-to-satellite tracking and on-board accelerometer data over a three months time span. For the first time it becomes possible to recover the gravity field from one satellite only. Thanks to CHAMP’s tailored orbit characteristics and dedicated instrumentation, providing continuous tracking and on-orbit measurements of non-gravitational satellite accelerations, the three months CHAMP-only solution provides the geoid and gravity with an accuracy of 20 cm and 1 mgal, respectively, at a half wavelength resolution of 550 km, which is already an improvement by a factor of two compared to any pre-CHAMP satellite-only gravity field model.

1. Introduction

With the launch of the German satellite CHAMP (Reigber *et al.*, 1999) on 15 July 2000, a new era in Earth gravity field recovery from space began. High-low satellite-to-satellite tracking (SST) using the American Global Positioning System (GPS) and on-board accelerometry combined with a low-altitude and almost polar orbit made CHAMP the first satellite being especially designed for long- to medium-wavelength global gravity field mapping. The on-board NASA/JPL BlackJack GPS space receiver provides two-frequency code and carrier phase range information (Kuang *et al.*, 2001) from up to twelve high-flying GPS satellites simultaneously. The French STAR accelerometer (Touboul *et al.*, 1999), accommodated in the centre of mass of CHAMP, measures in three dimensions the non-gravitational satellite accelerations caused by air density, wind and radiation pressure variations (surface forces). The orientation of the accelerometer’s axes is measured with respect to a celestial reference system by two star cameras. Figure 1 shows a sketch of the satellite with the gravity relevant instrumentation, and gives the orbit and mission parameters. The mean flight altitude of CHAMP, being initially 454 km, decreased with an average rate of 2 km/month over the first two years of the mission.



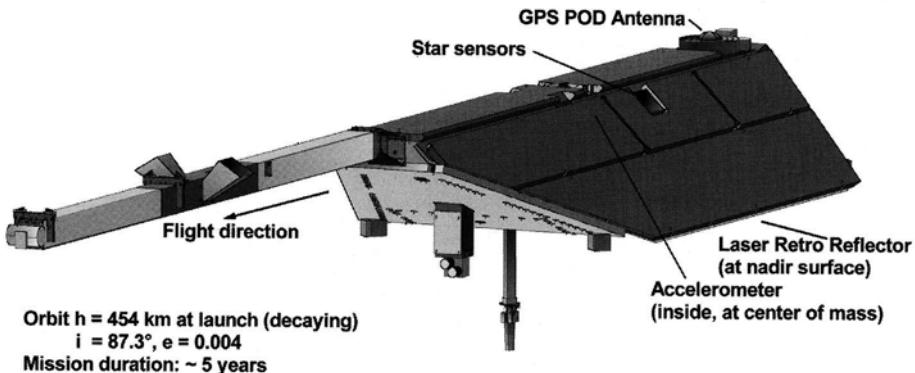


Figure 1. Sketch of the CHAMP satellite, orbit/gravity payload accommodation and mission parameters.

Compared to all former geodetic satellite missions used for global gravity field recovery, CHAMP has the following principal advantages: GPS SST yields a continuous multi-directional monitoring of the orbit compared to only one-dimensional sparse ground-based tracking during station overflights, and, being important for a very low flying satellite, the accelerometer measurements replace insufficient air drag modelling. By this, the purely gravitational orbit perturbation spectrum can be exploited for gravity field recovery all along the orbit (Figure 2) limited only by the instruments' noise and performance. In addition, the almost polar orbit of CHAMP provides a complete coverage of the near Earth environment with CHAMP observations.

It is the purpose of this contribution to demonstrate that with CHAMP it is for the first time possible to derive a global gravity field model based upon only one satellite and from only a few months' worth of tracking data. Moreover, the resulting model will prove to be superior in long-wavelength geoid and gravity field approximation to any pre-CHAMP satellite-only gravity field model like GRIM5-S1 (Biancale *et al.*, 2000) or EGM96S (Lemoine *et al.*, 1998), which were derived from multi-year tracking records of some tens of satellites. The gravity field model described in the following is called EIGEN-1 (European Improved Gravity Field Model of the Earth by New Techniques). Three months of CHAMP data out of July to December 2000 were processed to get the EIGEN-1 model.

2. Data Base, Data Processing and Solution Strategy

Global gravity field recovery from satellite orbit perturbations relies on a precise dynamic restitution of the satellite's trajectory (Reigber, 1989). Dynamic precise orbit determination means the numerical integration of the satellite's accelerations starting from a given state vector (position and velocity) at epoch T_0 . The accelerations are computed at each integration point from apriori known gravitational

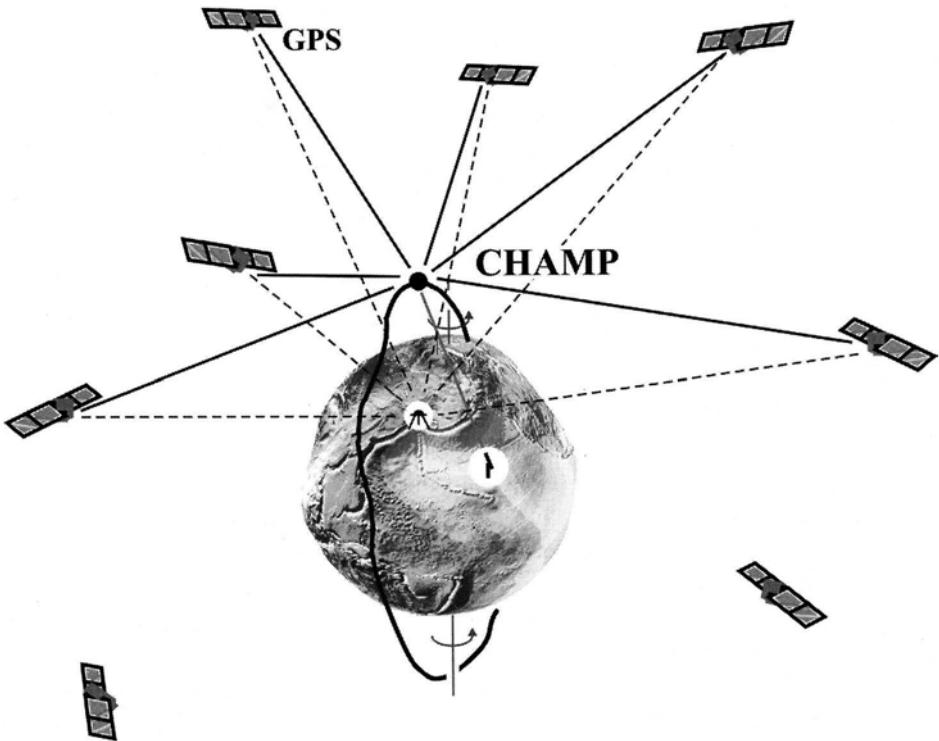


Figure 2. High-low GPS-CHAMP satellite-to-satellite tracking for gravity field recovery.

force field models (gravity including tidal and atmospheric temporal field variations) as well as, in the case of CHAMP, taken from the on-board accelerometer to account for non-gravitational perturbations. The integrated orbit (over one arc) is then fitted to the tracking observations (GPS-SST) in a least squares adjustment process solving for the so-called arc-dependent parameters, *i.e.* the state vector at epoch T_0 , and observation- and configuration-specific parameters as described below. For the determination of CHAMP orbits, a nominal arc-length of 1.5 days was selected. After convergence of the orbit adjustment with the a priori force field models, the observation equations are extended by partial derivatives for the so-called global unknowns, here the looked-for parameters describing the gravitational geopotential. Least squares equation systems are then generated arc-by-arc from the observation equations, accumulated to one overall system, which is eventually solved by matrix inversion. The nominal arc-length of 1.5 days was selected as a compromise between ‘short’, to avoid the increase of data and modelling errors during integration, and ‘long’, to retain resonant longer-period gravitational orbit perturbations.

For the EIGEN-1 gravity field solution 88 days of CHAMP data were processed covering the periods 2000, July 30 to August 10, and September 24 to December 31, with some gaps in between due to mission events and system unavailability.

The processing was split into 50 arcs of 1.5 days length and 13 arcs of 1 day length using the following data received from CHAMP:

GPS-CHAMP satellite-to-satellite tracking (SST) data. A total of 1.135 million code and carrier phase observations, respectively, were exploited, sampled at 30 s epochs from the originally received 10 s sequence. A maximum of (at that time) seven GPS satellites could be observed simultaneously at each epoch by the BlackJack space receiver. The receiver's clock offset per epoch with respect to the GPS sender clocks, and phase ambiguities per GPS satellite pass (about 675 unknowns per arc) have to be estimated in the data reduction process. There, the code and phase observations are weighted according to standard deviations of 22 cm and 1 cm, respectively.

Surface force accelerations. The along-, across-track and radial accelerations delivered with a frequency of 1 Hz by the STAR accelerometer were edited (outliers rejection) and compressed to 10 s normal points (mean over one 10 s interval). The 10 s interval coincides with the orbit integration step size. The resolution of the measured accelerations is $3 \times 10^{-9} \text{ m/s}^2$ for both the along- and across-track axis and one order of magnitude less for the radial axis. The calibration parameters (bias and scale factor) for each of the three accelerometer channels are estimated as unknowns simultaneously with the orbit and gravity field parameters. It turned out that the radial channel of the accelerometer is very sensitive to temperature variations due to a hardware failure. However, the radial channel is less important compared to the along- and across-track components, and also, the non-nominal behaviour could partly be repaired by solving for additional temperature-dependent calibration parameters.

Attitude quaternions. The orientation of the accelerometer's axes in space must be known with an accuracy of about 1 arcmin. The angles with respect to the celestial star catalog system are provided in terms of quaternions by two star camera heads which are rigidly connected with the accelerometer's cage. Like the accelerometer data, the 1 Hz sequence of the quaternions is edited and compressed to 10 s points.

Attitude thruster events. The attitude of CHAMP is controlled by six couples of cold gas thrusters which due to unknown residual mismatches not only rotate the satellite but also induce small linear accelerations each time one thruster pair is activated (about 10 times per revolution for 0.1 s to 3 s each). These residual accelerations are treated as solve-for parameters during data reduction.

The precise ephemerides and clock parameters of the 28 GPS satellites, needed when evaluating GPS-CHAMP SST data, are determined before-hand using GPS tracking data from the IGS (International GPS Service) ground station network

and CHAMP mission stations (Galas *et al.*, 2001). The pre-adjusted GPS positions and clocks then are held fixed in the subsequent CHAMP orbit computation and generation of gravity field normal equation systems.

The gravitational geopotential is mathematically described in the spectral domain by a spherical harmonic expansion (Heiskanen and Moritz, 1967). The Stokes' coefficients $\bar{C}_{l,m}$ and $\bar{S}_{l,m}$ of degree l and order m are the solve-for parameters. The selected solution space is complete to degree and order 91 with an increased resolution up to a maximum degree of 119 for the zonal coefficients and for terms within bands around order 16 and multiples of 16 up to the maximum order 111. These orders, close to m orbit revolutions per n days, are supposed to induce near-resonant, *i.e.* amplified and long-period orbit perturbations. The GM -term (gravitational constant G times mass of the Earth M) in the harmonic expansion was fixed in the adjustment because there is a one-to-one correlation with the accelerometer's bias parameters. Also, because of the short observation interval, no drift coefficients for the low degree zonals were estimated. The total number of unknown $\bar{C}_{l,m}$ and $\bar{S}_{l,m}$ coefficients amounts to 10738. Another 96 unknowns are associated with the periodic gravitational field variations of 8 diurnal and semi-diurnal ocean tidal waves.

Before solving the finally accumulated normal equation system it has to be stabilized because of a lack in sensitivity of CHAMP to higher degree and order terms whose signal at CHAMP's altitude gradually fades out. The stabilization constrains underdetermined harmonic coefficients to smaller values approaching zero if there is no information at all in the data.

After stabilization the normal equation system is solved by matrix inversion to yield the EIGEN-1 CHAMP-only global gravity field model and the standard deviations of the estimated parameters.

3. CHAMP-only Global Gravity Field Model EIGEN-1 and Solution Characteristics

Figure 3 shows the EIGEN-1 global gravity field model in terms of geoid heights as deduced from the adjusted spherical harmonic coefficients and related to a conventional ellipsoid of revolution. The power spectrum of the geoid, *i.e.* the signal amplitudes per degree of the spherical harmonic expansion, is shown in Figure 4. For comparison, the power spectra of the pre-CHAMP satellite-only model GRIM5-S1 and of the pre-CHAMP combined model EGM96 (Lemoine *et al.*, 1998) are also given. Both models have no data in common with the EIGEN-1 solution. EGM96 represents the overall geoid power as this model incorporates also terrestrial surface gravity and altimeter data, whereas GRIM5-S1 represents the pre-CHAMP state-of-art global gravity field processing from satellite tracking data only. The GRIM5-S1 data base covers multi-year tracking records from 24 satellites, compared to only three months of single satellite data used for the EIGEN-1 solution.

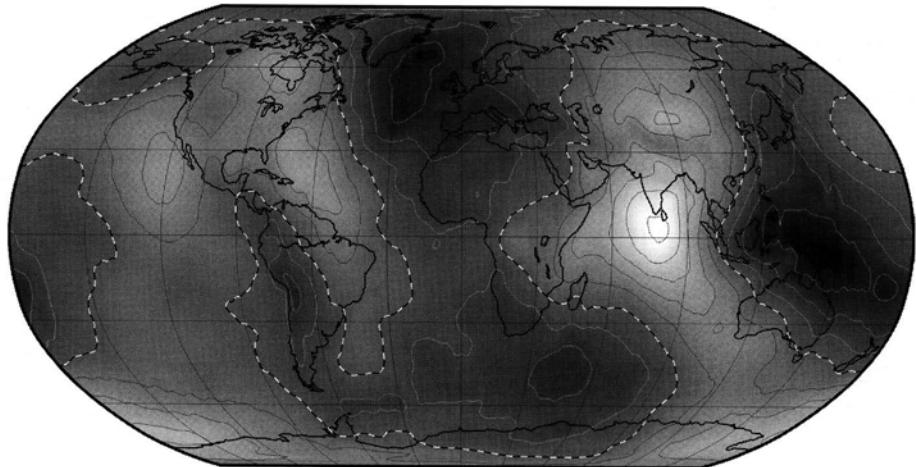


Figure 3. Geoid model EIGEN-1, shading from -100 m (white) to $+100$ m (black), contour interval: 20 m, black/white isolines: negative/positive values.

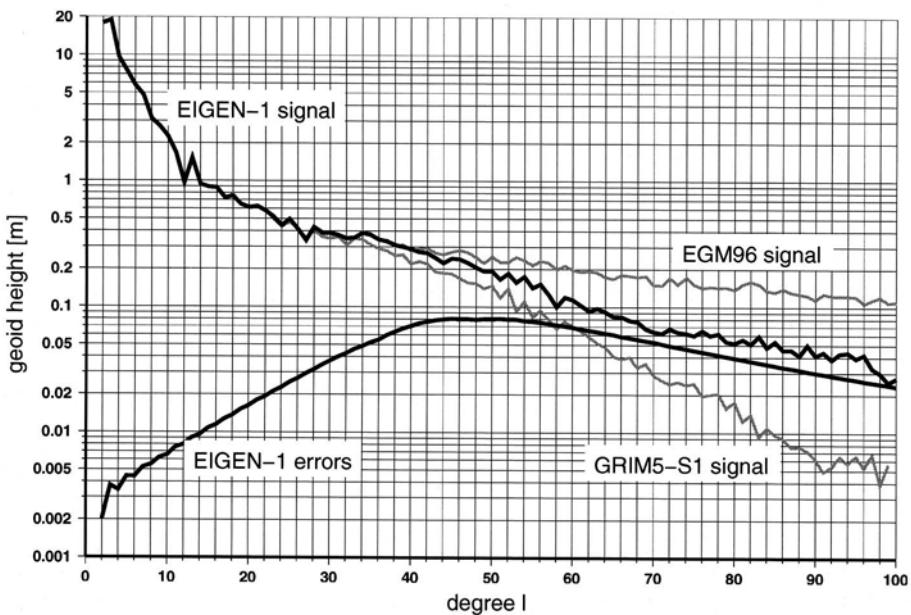


Figure 4. Signal/error amplitudes per degree in terms of geoid heights.

It can be seen from Figure 4, that EIGEN-1 resolves the geoid equally well like the other two models up to about degree 32 despite the large difference in the underlying data. As expected for higher degree terms the power of the two satellite-only solutions decreases with respect to the power of the combination solution EGM96 due to the signal attenuation with satellite altitude. Figure 4 also gives the error amplitudes per degree as computed from the coefficients' standard deviations

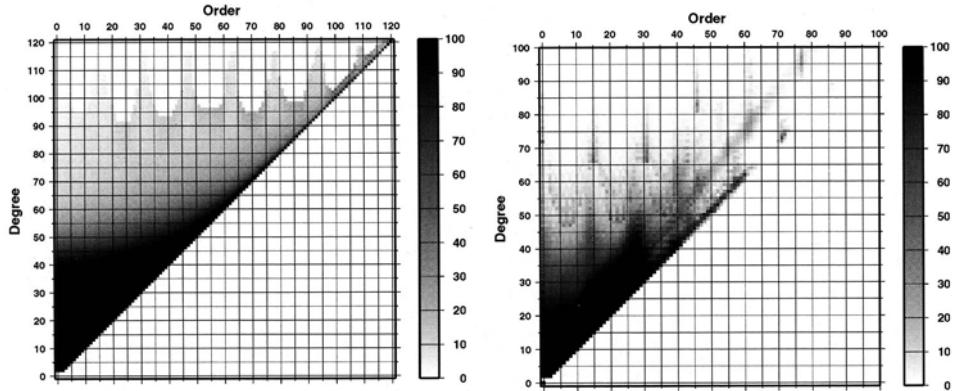


Figure 5. EIGEN-1 (left) and GRIM5-S1 (right) sensitivity matrices per degree and order of spherical harmonic coefficient (100%: fully resolved, 0%: not sensed at all by the satellites).

resulting from the adjustment. It turns out that the signal-to-noise ratio for the EIGEN-1 adjusted coefficients (root sum of squares over order per degree) gets smaller than one above degree 44. A complete resolution of the EIGEN-1 gravity field up to degree 32 in the spectral domain means a spatial resolution of 600 km (half wavelength) at the Earth's surface (Figure 3).

Another more detailed way of looking at the resolution, which is achieved with the CHAMP tracking data, is depicted in Figure 5. There, two sensitivity matrices are given, one for the GRIM5-S1 and one for the EIGEN-1 model. The shading indicates the contribution in per cent of the satellite tracking data to the determination of a particular solved-for spherical harmonic coefficient of given degree and order (either C or S coefficient) in relation to the pseudo-observation (value 0) introduced for matrix stabilization (Schwintzer, 1990). Again one realizes that both models resolve the gravitational spectrum for all terms up to about degree/order 32, but there is one remarkable difference between the two matrices: The GRIM5-S1 matrix reveals that the coefficients around orders 15 and multiples of 15, which at low altitudes induce resonant orbit perturbations are resolved to higher degrees than within non-resonant orders. This order specific pattern is completely absent in the EIGEN-1 sensitivity matrix which reveals a homogeneous swallow-tail pattern with a tendency of a higher degree resolution towards the sectorial terms, *i.e.* the coefficients of equal degree and order.

The explanation for the different structure of both normal equation systems is the uninterrupted coverage of CHAMP's orbit with multi-directional GPS satellite-to-satellite tracking data in the EIGEN-1 solution compared to sparse and pass-dependent ground-based tracking underlying the GRIM5-S1 solution. The continuous data coverage along the orbit and the complete coverage of the Earth's surface with ground tracks thanks to CHAMP's polar orbit result in a homogeneously conditioned normal equation system where the limitation in determinability

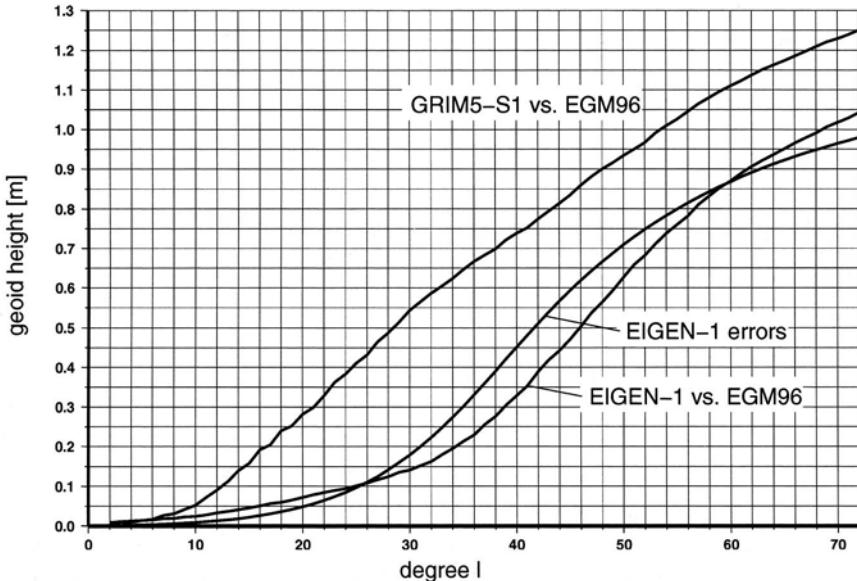


Figure 6. Difference/error amplitudes as a function of maximum degree in terms of geoid heights.

of the gravitational unknowns is only given by the orbit characteristics (and data noise) and not by the tracking data and ground station distribution.

As a consequence, the EIGEN-1 normal equation system needs stabilization only for terms with a degree larger than 29 whereas in the GRIM5-S1 solution all terms above degree 5 had to be stabilized for obtaining a reasonable solution.

4. Accuracy Evaluation of EIGEN-1

To get an insight into the accuracy of the CHAMP-only global gravity field solution, the differences of the spherical harmonic coefficients between EIGEN-1 and the completely independent model EGM96 are computed. EGM96 is taken as a reference because it also contains high-resolution surface data from gravimetry and altimetry. Figure 6 gives the difference amplitudes in terms of geoid heights as a function of maximum degree l_{\max} (*i.e.* root sum of squares of all values with degree smaller and equal l_{\max}). For comparison, Figure 6 also contains the corresponding curve obtained when subtracting GRIM5-S1 from EGM96. The gain in accuracy when using CHAMP data in the recovery of the long-wavelength geopotential is striking: an improvement of the agreement with EGM96 by a factor of four up to degree/order 20 (GRIM5-S1: 27 cm, EIGEN-1: 7 cm) and a factor of three up to degree/order 30 (55 cm and 18 cm, respectively). Under the assumption that, when two independent models agree better, then both are closer to reality, one can conclude that the improvement in the agreement with EGM96 corresponds to a real accuracy improvement of EIGEN-1 with respect to GRIM5-S1.

TABLE I

Gravity Field Model Comparisons with Altimetric Ocean Geoid Heights (N) and Oceanic Gravity Anomalies (Δg); Weighted (Cos of latitude) Root Mean Square of Differences About Mean.

Grid Spacing	EIGEN-1 / GRIM5-S1 / EGM96S	
	N	Δg
5 × 5 degree	22 / 41 / 35 cm	1.3 / 2.0 / 1.8 mgal
2.5 × 2.5 degree	49 / 69 / 64 cm	5.0 / 5.4 / 5.4 mgal

The comparison with EGM96 gives a conservative estimate of EIGEN-1's accuracy over the gravitational spectrum because the differences also contain the errors of EGM96. From Figure 6 one can deduce that a geoid accuracy of 10 cm is achieved by EIGEN-1 for a spectral resolution of degree/order 26, corresponding to a spatial resolution of 760 km (half-wavelength), then decreasing to an accuracy of 22 cm for a degree/order 36 or 550 km resolving geoid, *e.g.* represented by 5 × 5 degree block mean values. Compared to the pre-CHAMP satellite-only model GRIM5-S1 the geoid resolution for the two given accuracy levels could be doubled. The EIGEN-1's error amplitudes as a function of maximum degree in Figure 6 represent the spectrum of the formal standard deviations as obtained from the adjustment. The curve partly exceeds the difference amplitudes proving that the estimated standard deviations are partly too pessimistic.

To further evaluate the quality of the EIGEN-1 gravity field model, comparisons are made in the spatial domain with homogeneous altimeter derived ocean surface data in terms of geoid heights and gravity anomalies. For this purpose, geoid heights and gravity anomalies, both as block mean values over a regular grid with a spacing of 5 degree as well as 2.5 degree, were computed form the EIGEN-1, GRIM5-S1 and the EGM96S Stokes gravitational coefficients. EGM96S, like GRIM5-S1, is a pre-CHAMP satellite-only model underlying the US EGM96 combined solution. These model data sets were then compared with (1) ocean geoid heights from ERS-/Topex altimetry (GFZ internal solution) after correcting for the sea surface topography according to the POCM model (Semptner and Chervin, 1992) and (2) oceanic gravity anomalies (Lemoine *et al.*, 1998), both data sets averaged to represent a 5 degree and 2.5 degree grid resolution and filtered to eliminate higher frequency content (above 5 degree and 2.5 degree half wavelength, respectively). The root mean squares of the differences (after bias elimination) are given in Table I, showing the very substantial improvement of the EIGEN-1 model compared to the pre-CHAMP satellite-only models GRIM5-S1 and EGM96S.

Table I confirms the 20 cm EIGEN-1 geoid accuracy at a spatial resolution of 5 × 5 degree, corresponding to a spectral resolution of degree/order 36. The

gravity error for 5×5 degree EIGEN-1 derived block mean values amounts to 1 mgal, when reducing the value in Table I by the error in the comparison data. For the 2.5×2.5 degree grid spacing, which corresponds to a spectral resolution of degree/order 72, the additional omission error due to the loss of power for higher degree terms in the satellite-only models becomes visible.

The performance of a new global gravity field model in dynamic precise orbit determination of a number of representative satellites usually is taken as a criterion to describe the quality of a gravity field model. As expected, the orbital fit of CHAMP GPS tracking data (carrier phases) improves from 20 cm to 1 cm and of laser tracking data from about 1 m to 10 cm, when replacing GRIM5-S1 by EIGEN-1 in the integration and adjustment of CHAMP's orbit with data outside the gravity field processing period. The opposite behaviour can be observed for other low Earth orbiting satellites like GFZ-1, Starlette, Stella and ERS. Because tracking data from these satellite were extensively exploited in the GRIM5-S1 gravity field model, especially the higher degree terms within the orbit specific resonant orders are well resolved in GRIM5-S1 and not well covered by the single satellite model EIGEN-1 as can be seen in Figure 5. Therefore, for the purpose of general precise satellite orbit determination, a combination of the EIGEN-1 and GRIM5-S1 normal equation systems is necessary. The resulting model EIGEN-1S is described in Reigber *et al.* (2002).

5. Conclusions

The exploitation of three month's worth of CHAMP data has proven that the combination of nearly continuous high-low satellite-to-satellite tracking and accelerometry aboard a satellite in a low altitude and highly inclined orbit allows a drastic improvement in long-wavelength global gravity field recovery. The CHAMP-based gravity field model EIGEN-1 provides the geoid and gravity with an accuracy of 20 cm and 1 mgal, respectively, at half wavelengths down to 550 km, which is an improvement by a factor of two compared to pre-CHAMP satellite-only models. For longer wavelengths a gain in accuracy by even a factor of four could be obtained. Oceanography and geodesy will largely benefit from this progress when applying the geoid as a dynamic reference surface.

It became for the first time possible to recover the global gravity field with data from one satellite only and from only a few months observation time. The high quality, strength and homogeneity of the long-wavelength geopotential solution will also allow to resolve temporal field variations at longest wavelengths. For precise orbit determination of low-flying satellites in different orbits than CHAMP a combination with normals from tracking data of such satellites is still required.

CHAMP has been the forerunner of a new generation of dedicated gravity satellite missions, followed by the twin satellite mission GRACE (Tapley and Reigber, 2001), launched in March 2002, with its low-low satellite-to-satellite link aiming at

a high resolution of temporal field variations, and ESA's mission GOCE (European Space Agency (ESA, 1999) scheduled for 2006 with on-board gradiometry so as to obtain an ultimately high spatial resolution and accuracy.

Acknowledgements

The German Ministry of Education and Research (BMBF) supported the CHAMP Phase C/D project under grant 50EP9087 and supports the exploitation phase under grant 50EP9587. On the French side the research was supported by the Centre National d'Etudes Spatiales (CNES).

References

- Biancale, R., Balmino, G., Lemoine, J.-M., Marty, J.-C., Moynot, B., Barlier, F., Exertier, P., Laurain, O., Gegout, P., Schwintzer, P., Reigber, C., Bode, A., König, R., Massmann, F.-H., Raimondo, J.-C., Schmidt, R. and Zhu, S. Y.: 2000, 'A new global Earth's gravity field model from satellite orbit perturbations: GRIM5-S1', *Geophys. Res. Letters* **27**, 3611–3614.
- European Space Agency: 1999, 'Gravity Field and Steady-State Ocean Circulation Mission (GOCE)', *The four candidate Earth explorer core missions*, Report for mission selection, Noordwijk, The Netherlands, SP-1233 (1), 217 pp.
- Galas, R., Wickert, J. and Burghardt, W.: 2001, 'H/R low latency GPS ground tracking network for CHAMP', *Phys. Chem. Earth (A)* **26**, 619–652.
- Heiskanen, W. and Moritz, H.: 1967, *Physical Geodesy*, W.H. Freeman, San Francisco, 364 pp.
- Kuang, D., Bar-Server, Y., Bertiger, W., Desai, S., Haines, B., Iijima, B., Kruizinga, G., Meehan, Th. and Romans, L.: 2001, 'Precise orbit determination for CHAMP using GPS data from BlackJack receiver', *2001 ION National Technical Meeting Proceedings, Session E1: Scientific Applications, Timing, and Frequency*, Long Beach, California, 9 pp.
- Lemoine, F. G., Kenyon, S. C., Factor, J. K., Trimmer, R. G., Pavlis, N. K., Chinn, D. S., Cox, C. M., Klosko, S. M., Luthcke, S. B., Torrence, M. H., Wang, Y. M., Williamson, R. G., Pavlis, E. C., Rapp, R. H. and Olson, T. R.: 1998, 'The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96', NASA Technical Paper, Goddard Space Flight Center, Greenbelt, NASA/TP- 1998-206861, 509 pp.
- Reigber, C.: 1989, 'Gravity field recovery from satellite tracking data', in F. Sansò, R. Rummel (eds.), *Theory of satellite geodesy and gravity field determination, Lecture notes in Earth sciences* **25**, Springer, Berlin, Heidelberg, New York, 197–234.
- Reigber, C., Schwintzer, P. and Lühr, H.: 1999, 'The CHAMP geopotential mission', *Boll. Geof. Teor. Appl.* **40**, 285–289.
- Reigber, C., Balmino, G., Schwintzer, P., Biancale, R., Bode, A., Lemoine, J.-M., König, R., Loyer, S., Neumayer, H., Marty, J.-C., Barthelmes, F., Perosanz, F. and Zhu, S. Y.: 2002, 'A high-quality global gravity field model from CHAMP GPS tracking data and accelerometry (EIGEN-1S)', *Geophys. Res. Lett.* **29** (14), 10.1029/2002GL015064.
- Schwintzer, P.: 1990, 'Sensitivity analysis in least squares gravity field modelling by means of redundancy decomposition of stochastic a-priori information', Internal report PS/51/90/DGFI, DGFI Dept. 1, Munich, 14 pp.
- Semptner, A. J. and Chervin, R. M.: 1992, 'Ocean circulation from a global eddy-resolving model', *J. Geophys. Res.* **97** C4, 5493–5550.

- Tapley, B. D. and Reigber, C.: 2001, 'The GRACE mission: status and future plans', *Eos Trans AGU* **82** (47), Fall Meet. Suppl., G41 C-02.
- Touboul, P., Willeminot, E., Foulon, B. and Josselin, V.: 1999, 'Acclerometers for CHAMP, GRACE and GOCE space missions: synergy and evolution', *Boll. Geof. Teor. Appl.* **40**, 321–327.

Address for Offprints: Peter Schwintzer, GeoForschungsZentrum Potsdam, Dept. 1, Telegrafenberg A 17, 14473 Potsdam, Germany (e-mail: psch@gfz-potsdam.de)

THE PROCESSING OF BAND-LIMITED MEASUREMENTS; FILTERING TECHNIQUES IN THE LEAST SQUARES CONTEXT AND IN THE PRESENCE OF DATA GAPS

W.-D. SCHUH (schuh@uni-bonn.de)

Institute of Theoretical Geodesy, University of Bonn, Nussallee 17, D-53115 Bonn, Germany

Received: 15 May 2002; Accepted in final form: 20 August 2002

Abstract. This paper discusses the treatment of correlated measurements in the least squares context. We focus on the processing of band-limited measurements and on long time series with a constant sampling interval. Time domain as well as frequency domain approaches were discussed to offer different ways to integrate the filtering process into the optimization scheme as good as possible. The focus was on long equispaced data sets. The application of discrete filters in the space domain makes it possible to decorrelate the observations during data acquisition. This opens the way to a sequential adjustment procedure, where the design matrix is treated row-by-row. Huge systems with millions of observations can be solved by direct or iterative strategies, and both approaches benefit from well-tailored filter techniques. Because of the sequential access the computational effort of this giant task can be easily distributed to a cluster of parallel processors and offers, in addition, the possibility to treat data gaps in a straightforward way.

Keywords: least squares, band-limited observations, decorrelation, filtering, whitening process

1. Introduction

Satellite gravity gradiometry (SGG) data enable the recovery of the detailed structures of the Earth's gravity field, which are represented by a set of the order of fifty to hundred thousands of parameters. This large number of parameters has to be derived from an even larger number of correlated observations (two six month measurement periods with a sampling rate of 1s). This inverse problem can be dealt with in a very efficient way based on an analytic representation by means of the inclination function (Schrama, 1991; Sneeuw, 2000). This approach allows it to handle the correlations of the observations in the frequency domain. On the other hand, strong restrictions must be imposed to perform this semi-analytical approach. In order to determine realistic measurement scenarios with perturbed orbits, satellite maneuvers, calibration phases and measurement interruptions, only a "brute force" least squares strategy (with the exact observation equations being solved) is able to cope with such kinds of situations. Apart from the huge numerical computational problem, the modeling of correlated observations is another difficulty, because it seems to be necessary to invert (solve) the variance/covariance matrix.



Space Science Reviews **108:** 67–78, 2003.

© 2003 Kluwer Academic Publishers.

In geodesy two different approaches are used to model correlated observations. In the space/time domain the autocovariance function allows a very efficient modeling of the correlations. Particularly for small data sets, an individual treatment of each observation is possible. An irregular data distribution and different types of data can be assimilated to a consistent model (Moritz, 1980). But in general, it is impossible for large data sets to handle the covariance matrix, because the dimension of the matrix grows with the number of observations. Therefore, it is necessary to introduce additional restrictions – equispaced data, periodicity – to get special types of matrices (Toeplitz, circulant matrices), which can be treated very efficiently.

Gathering all the measurements in an observation vector $\ell \in IR^{n \times 1}$ and the corresponding error covariances in the symmetric, positive definite matrix $\Sigma\{\ell\} \in IR^{n \times n}$ we use the Gauss-Markov model to solve the overdetermined system

$$\ell + v = Ax, \quad \Sigma\{\ell\} = \Sigma \quad (1)$$

where the design matrix $A \in IR^{n \times m}$ contains the linear (or linearized) relation between observations ℓ and the parameters $x \in IR^{m \times 1}$. Applying the least squares optimization principle

$$\langle v, v \rangle_{\Sigma^{-1}} = v^T \Sigma^{-1} v \dots \min \quad (2)$$

with respect to the Σ^{-1} norm leads to the normal equation system

$$(A^T \Sigma^{-1} A) \tilde{x} = A^T \Sigma^{-1} \ell \quad (3)$$

and to the best linear unbiased estimator \tilde{x} of the parameter x . With respect to gridded data sets or sequences of equispaced samples the covariance matrix Σ has a Toeplitz structure*. The spectral representation of the symmetric covariance matrix is given by

$$\Sigma = U \Lambda U^T \quad (4)$$

with the diagonal matrix $\Lambda \in IR^{n \times n}$ of eigenvalues and with the orthogonal matrix $U \in IR^{n \times n}$ of eigenvectors. Because of the regular structure these eigenvectors can be expressed in terms of cosine and sine functions and therefore the eigenvalues of the covariance matrices describe immediately the frequency behavior of the measurement series.

The measurement series is transferred to the frequency domain (Fourier transform), the power spectral density – the frequency domain representation of the autocovariance function – is used to model the behavior of the correlated observations. This frequency domain strategy requires to have the whole measurement

* Unfortunately, the inversion process destroys the regular structure, but in this application we deal only with periodic and long sequences of data streams, and therefore, we treat the Toeplitz systems like circulant or infinite systems.

series at once, so this series must be complete in the sense of equispaced data within the entire period (periodicity). Otherwise the transformation from the space domain into the frequency domain in a strict form is impossible. Due to the convolution theorem the time consuming circular convolution – a matrix-vector operation – in the space domain corresponds to an element-wise multiplication in the frequency domain. Therefore, the correlations of the observations, represented by the covariance matrix can be simply replaced by a multiplication frequency-by-frequency. This frequency domain approach is often denoted as fast collocation (Bottani and Barzaghi, 1993).

In the next section, we will introduce a decorrelation process to transform the measurement series into a form, where each measurement can be treated independently. Then we will summarize the characteristic quantities of a general linear process and introduce the decorrelation process into this terminology. We will see that the decorrelation process corresponds to the whitening procedure, where the target of the filter design is to transform a colored noise of a random process into a white Gaussian noise.

The use of discrete filters in the time domain allows it in particular to treat measurement series observation by observation. The advantage of this sequential access can be immediately utilized within the solution process. These computational aspects to treat huge least squares problems and to distribute the numerical effort to a cluster of computers is described in Sec. 5. It is shown that both, the direct and the iterative solver, benefit from the sequential access and that parallel processing strategies can be applied.

2. Decorrelation

The variance/covariance matrix Σ represents the accuracy as well as the correlations between observations. This symmetric, positive definite matrix Σ can be factorized into two regular, upper triangular systems R

$$\Sigma = R^T R, \quad R \in I\mathbb{R}^{n \times n}. \quad (5)$$

The inverse R^{-1} of the triangular system can be used to multiply the observation equations from the left and to substitute the Gauss-Markov model (1) by

$$\bar{\ell} + \bar{v} = \bar{A}\bar{x}, \quad \Sigma\{\bar{\ell}\} = \bar{\Sigma} \quad (6)$$

with

$$\bar{\ell} = (R^{-1})^T \ell, \quad \bar{v} = (R^{-1})^T v, \quad \bar{A} = (R^{-1})^T A. \quad (7)$$

The covariance matrix $\bar{\Sigma}$ of the transformed observations has to be determined by variance propagation. Introducing eqn. (5) we end up with

$$\bar{\Sigma} = (R^{-1})^T \Sigma R^{-1} = I, \quad (8)$$

where \mathbf{I} represents the identity matrix. This means that the observations $\bar{\ell}$ of the substitute model (6) are uncorrelated and that the optimization principle is reduced to

$$\langle \bar{v}, \bar{v} \rangle_{\mathbf{I}} = \bar{v}^T \bar{v} \dots \min . \quad (9)$$

If we denote the matrix $(\mathbf{R}^{-1})^T$ by \mathbf{F}

$$\mathbf{F} = (\mathbf{R}^{-1})^T \quad (10)$$

the transformation (7) can be rewritten as

$$\bar{\ell} = \mathbf{F}\ell, \quad \bar{v} = \mathbf{F}v, \quad \bar{A} = \mathbf{F}\mathbf{A} . \quad (11)$$

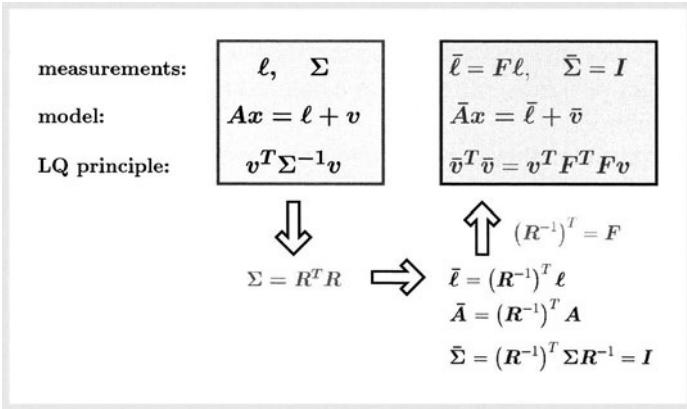


Figure 1. Decorrelation Process.

This decorrelation process, which is summarized in Fig. 1, works for any covariance matrix, but if the covariance matrix possesses a special structure, e.g., circulant structure because of periodic equispaced data, the factorized matrices \mathbf{R} and \mathbf{F} will reflect this property as well (pay regard to footnote on page 2), and we end up with triangular Toeplitz systems. The multiplication of a Toeplitz system with a vector (11) can be interpreted as a convolution of the first column of \mathbf{F} , denoted by f , with the vector ℓ . Thus we obtain

$$\bar{\ell} = f * \ell , \quad \bar{\ell}(n) = \sum_{k=1}^n f(k) \ell(n-k) \quad (12)$$

or

$$\bar{a}^{(j)} = f * a^{(j)} , \quad j = 1, \dots, m , \quad (13)$$

respectively, for each column $a^{(j)}$ of the design matrix \mathbf{A} . This convolution can be seen as the application of a discrete, linear, shift-invariant filter, and the vector f represents the impulse response of the filter.

3. General Linear Process

A *general linear process* is given by the convolution

$$y = x * h \quad (14)$$

which may be written in sequence form

$$y(n) = x(n) * h(n) = \sum_{k=-\infty}^{\infty} h(k) x(n-k) \quad (15)$$

of a stationary random sequence $x(n)$ by a linear shift-invariant filter defined by the impulse response $h(n)$. The mean of this process

$$m_y(n) = E\{y(n)\} = E \left\{ \sum_{k=-\infty}^{\infty} h(k) x(n-k) \right\} = m_x \sum_{k=-\infty}^{\infty} h(k) = m_y \quad (16)$$

is independent of the time n , and thus the autocorrelation sequence

$$r_{yy}(k) = r_{xx}(k) * h(k) * h(-k) \quad (17)$$

is also shift-invariant, and depends only on time differences expressed by k . If we consider processes with zero mean $m_x = 0$, the autocorrelation sequences $r_{xx}(k)$ and $r_{yy}(k)$ as second moments correspond to the autocovariance sequences $\sigma_{xx}(k)$ and $\sigma_{yy}(k)$ as second central moments

$$\sigma_{yy}(k) = \sigma_{xx}(k) * h(k) * h(-k). \quad (18)$$

The power spectral density of the filtered process is given by

$$P_{yy}(e^{j\omega}) = P_{xx}(e^{j\omega}) |H(e^{j\omega})|^2 \quad (19)$$

where $H(e^{j\omega})$ denotes the frequency response of the filter defined by the Fourier transformed impulse response sequence $h(n)$

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k) e^{-jk\omega} = \mathcal{F}\{h(k)\}. \quad (20)$$

The power spectral densities P_{yy} and P_{xx} are obtained by the Wiener-Khinchin theorem as the Fourier transform of the autocorrelation sequence $r_{yy}(k)$

$$P_{yy}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_{yy} e^{-jk\omega} = \mathcal{F}\{r_{yy}(k)\} \quad (21)$$

and $r_{xx}(k)$, respectively. In terms of z -transforms eqn. (19) is written as

$$P_{yy}(z) = P_{xx}(z) H(z) H^*(1/z^*). \quad (22)$$

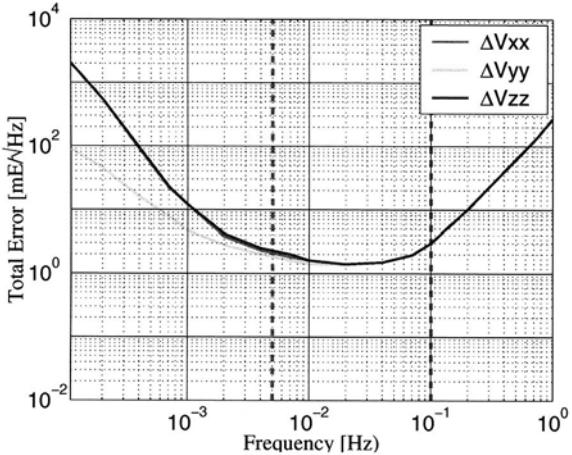


Figure 2. Square root of the spectra of the gravity-gradient measurement error budget for the components ΔV_{xx} , ΔV_{yy} , and ΔV_{zz} (*Granada-SGG* error budget, cf. ESA (1999), pp. 177, digitized fig. 8.2). The error level in the measurement bandwidth between 5×10^{-3} Hz and 0.1 Hz is below $\pm 3 \text{ mE}/\sqrt{\text{Hz}}$.

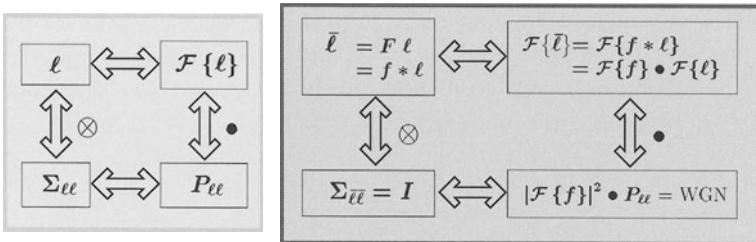


Figure 3. Filter Design - PowerSpectralDensity.

where z can be interpreted as a shift operator $z^\alpha h(k) = h(k - \alpha)$. The reverse sequence $h(-k)$ of the impulse response $h(k)$ is replaced by the conjugated reciprocal form $H^*(1/z^*)$.

Thus, if we use a white noise process as input process $x(n)$, the frequency response function $H(e^{j\omega})$ of the filter will determine the spectral behavior of the output process $y(n)$. This relation is used to generate colored sequences of measurement noise. In our simulations of the performance of the GOCE gradiometer mission, we start with a given power spectral density function, which was deduced from simulation studies and describes the spectral behavior of the gradiometer measurements (cf. Fig. 2). We compute the square root form and design a linear shift-invariant filter with a frequency response function, which reflects the same behavior.

The measurements are decorrelated by an inverse procedure. We start with a colored input sequence $x(n)$ and try to find a filter, which “pre-whitens” this se-

quence. The resulting sequence $y(n)$ has to reflect white Gaussian noise (WGN), which means that the power spectrum density function does not favor particular frequencies. The flattening of P_{yy} may be obtained by multiplying the power spectral density P_{xx} of the colored sequence $x(n)$ by a function $|\mathcal{F}\{f\}|^2$, which reflects the inverse behavior of the power spectral density of the measurements (cf. Fig. 2).

In general, a linear shift-invariant discrete system (filter) can be represented by the difference equation

$$y(n) + \sum_{l=1}^{n_a} a_l y(n-l) = \sum_{l=0}^{n_b} b_l x(n-l) \quad (23)$$

with a finite number n_a and n_b of constant coefficients a_l and b_l . The corresponding frequency response function (denoted here as filter *transfer* or *system function*) is given by

$$H(z) = \frac{\sum_{l=0}^{n_b} b_l z^{-l}}{1 + \sum_{l=1}^{n_a} a_l z^{-l}}. \quad (24)$$

Signal processing tools allow to design discrete linear filters, which are able to express the correlation between the observations. In general, there are two types of filters:

- *Nonrecursive filters*, describing the relation between input and output sequence by a convolution. These filters are also referred to as *convolution* or *transversal filters*. They are represented by a moving average operation in the time domain and by the product of the spectra as given by (19) in the frequency domain.
- *Recursive filters*, for which the filter output $y(n)$ at time $t_n = n$ is computed recursively from the filter output $y(j)$, $j < n$ and the input sequence $x(k)$, $k \leq n$.

Therefore, the summation indices n_a and n_b in eqn. (23) determine the type of the system, which is often denoted as Auto-Regressive-Moving-Average (ARMA) filter

- $n_a = 0, n_b > 0 \dots$ convolution filter, ARMA(0, n_b) filter,
- $n_a > 1, n_b = 0 \dots$ recursive filter, ARMA(n_a ,0) filter and
- $n_a > 1, n_b > 0 \dots$ ARMA(n_a, n_b) filter.

All types of filters have their advantages and disadvantages. The examples in Sec. 4 show two typical applications of pre-whitening filters in least squares problems, where the recursive filter ARMA(2,2) was applied in order to minimize the number of operations during the assembling of the normal equations. In contrast, a moving average filter ARMA(0,88) is implemented in the iterative solution process, because the symmetry of this type of filters allows an efficient access to compute the steps of an iterative scheme (cf. Sec. 5).

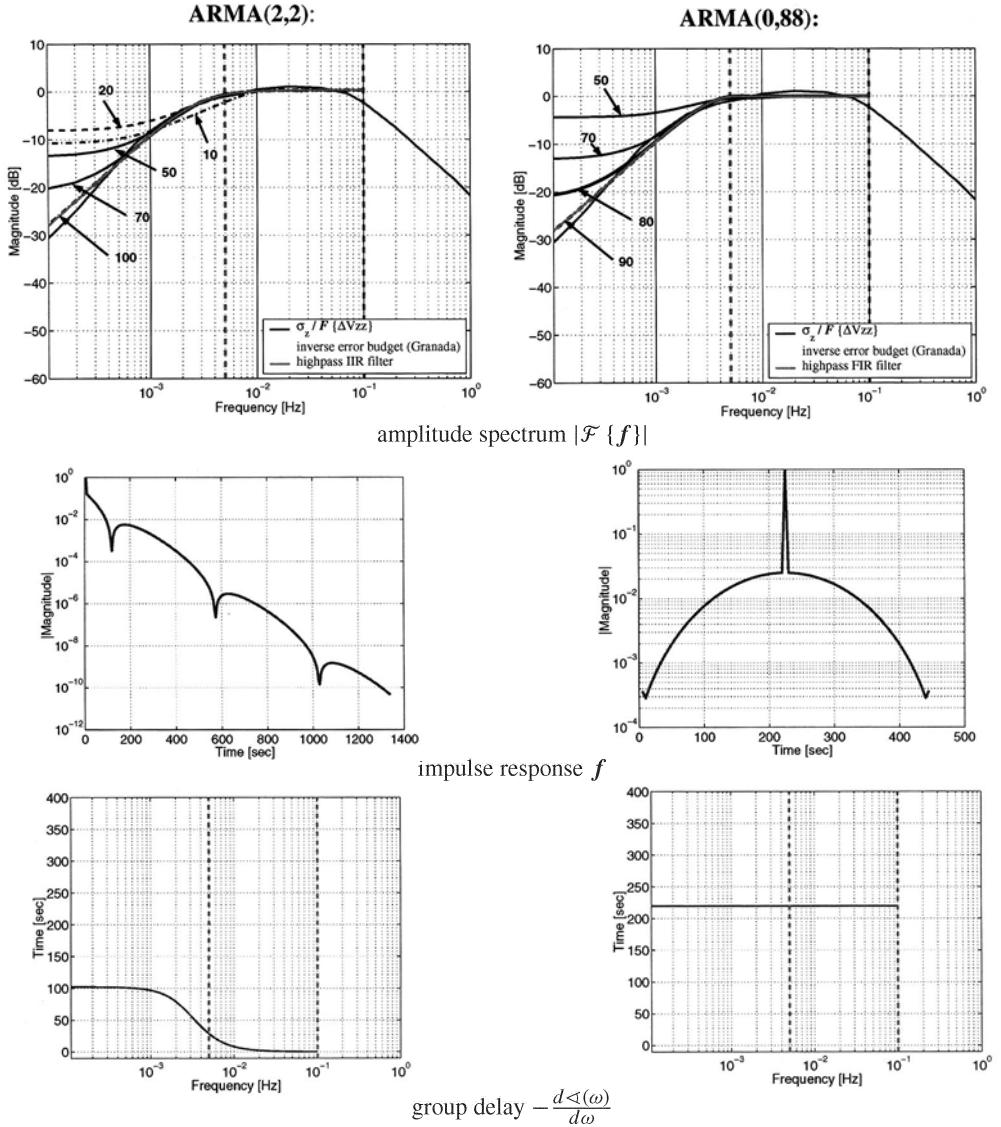


Figure 4. Characteristics of an AutoRegressiveMovingAverage filter of the order 2 (ARMA(2,2)) and a MovingAverage filter of the order 88 (ARMA(0,88)).

4. Characteristics of the Discrete Filters

As an example we have designed two linear discrete filter sets capable of working as a pre-whitening filter for gradiometry measurements (e.g., component ΔV_{zz}) with a power spectral density function given by Fig. 2.

Fig. 4 summarizes the characteristics of an AutoRegressiveMovingAverage filter of the order 2 (ARMA(2,2)) and a MovingAverage filter of the order 88

(ARMA(0,88)). On top of Fig. 4 the amplitude spectra reflect the spectral behavior with respect to the inverse error budget of the gradiometer measurements (component ΔV_{zz}). Special attention is paid to the warm-up phase of the filter. The curves after 10, 20, ..., 90 samples reflect the boundary condition and, therefore, the fill up of the intern memory of the filter-box. In the middle of Fig. 4 the impulse response of the filters are drawn. In contrast to the infinite, slow decreasing impulse response of the ARMA(2,2) the finite impulse response of the ARMA (0,88) filter reflects a symmetric behavior.

The direct access to the behavior in boundary regions (data gaps) is a nice property of the finite impulse response filter. The warm-up phase at the very beginning of a sequence or in case of a data gap can be immediately deduced from the support of the finite impulse response. For shorter time periods the spectral behavior is given by the Fourier transform of the truncated impulse response. Fig. 4 shows this warm-up behavior. The amplitude spectra for both mentioned filters indicates a warm-up time of about 90 samples (450 s) to reproduce the spectral behavior of the gradiometer measurements. At the beginning low frequencies are disturbed in particular and, therefore, a warm-up phase is necessary to fill the memories of the filter-box appropriately.

At bottom of Fig. 4 the deviation of the phase spectrum $\triangle(\omega)$ with respect to the frequency ω – the group delay – of the filter gives an impression of the delay in units of time with regard to individual frequencies. The left figure shows that different frequencies use varying times to be transported through the filter. For low frequencies it takes 100 s (20 samples) to penetrate the filter and high frequencies pass the filter without delay. In contrast to this frequency dependent delay the MA-filter (see right diagram) shows a constant delay (220 s, 44 samples) for all frequencies.

5. Computational Aspects

All types of filters can be applied in the time domain (23) or in the frequency domain, where the signal x is Fourier transformed and denoted as sample spectrum or periodogram $\mathcal{F}\{x\}$ and the impulse response function h is transformed to the frequency response function $\mathcal{F}\{h\}$. The convolution in the time domain is translated into a simple multiplication in the frequency domain (cf. Fig. 3). With regard to large adjustment problems, the time domain approach has the advantage that the filter process can be applied to data during acquisition. Therefore, a sequential adjustment procedure can be applied, which means that the contribution of each individual observation equation can be dealt with sequentially. The typical operations, the computation of the normal equations

$$N = \bar{A}^T \bar{A}, \quad n = \bar{A}^T \bar{\ell} \quad \text{or} \quad N = (\mathbf{F}\mathbf{A})^T \mathbf{F}\mathbf{A}, \quad n = (\mathbf{F}\mathbf{A})^T \mathbf{F}\ell \quad (25)$$

using the direct solver, and the determination of the form

$$q = \bar{A}^T \bar{A} p \quad \text{or} \quad q = \mathbf{A}^T \mathbf{F}^T \mathbf{F}\mathbf{A} p \quad (26)$$

```

! initialize filter array A
INIT_FILTER_A(A_row)
! initialize scalar filter ℓ
INIT_FILTER_ℓ(ℓ)
q = 0
n = 0

DO k = 1, ..., #n_obs
  [A_row, ℓ] = COMPUTE_OBS(k)
  ! filter A_row column wise
  Ā_row = FILTER_A(A_row)
  ē = FILTER_ℓ(ℓ)
  ! update N by a dyadic product
  N = N + Ā_rowT Ā_row
  ! update n by a scaled vector
  n = n + Ā_rowT ē
END DO k

```

```

! initialize scalar filter h
INIT_FILTER_h(h)
! initialize scalar filter ī
! scalar filter
INIT_FILTER_ī(h)
q = 0
DO k = 1, ..., #n_obs
  A_row = COMPUTE_OBS(k)
  ! scalar product
  h = A_row • p
  ! filter scalar h
  ī = FILTER_h(h)
  ! filter scalar ī
  ī = FILTER_ī(h)
  ! update q by a scaled vector
  q = q + A_rowT ī
END DO k

```

Algorithm 1. Efficient implementation of the filter for large design matrices A . The computational steps to compute the normal equation $N = (FA)^T(FA)$ shows the algorithm on the left-hand side. On the right-hand side an efficient algorithm to compute the form $q = A^T FFAp$ is presented. This form has to be computed during each iterative solution. The shape of the input vector p is used to reduce the computational steps by solving this product from the right to the left. In this case only scalar products, scalar filter operations and a scaled vector update are necessary. However, to preserve the sequential access to the observation equations, a symmetric filter operator $F^T = F$ is assumed.

with respect to iterative solvers can be performed row-by-row, using each row $A(k, :)$ of the design matrix A only once. This is important when dealing with large and very large systems, not allowing it to store the entire matrix A and requiring the computation and processing to be done row-by-row.

The filter in the time domain (23) can be seen as a box, in which (within each step) a new input value $x(n)$ produces one output value $y(n)$. The box stores the coefficients, as well as some of the past results in a special memory-stack. For details, confer Schuh (1996, App. B1), or Oppenheim and Schafer (1975, Sec. 4). Therefore, the operations $\ell = F\ell$ can be performed step-by-step by $\bar{\ell}(k) = \text{FILTER}(\ell(k))$. The same may be done with each column of the matrix A to compute the product $\bar{A} = FA$, where a row-wise access yields $\bar{A}(k, :) = \text{FILTER}(A(k, :))$. Here, an array of filter-boxes is applied. Each column possesses its own filter-box, using the same filter operators (coefficients), but each box is equipped with its own memory-stack (single instruction - multiple data). Each call of the filter operation produces a new result $\bar{A}(k, :)$ and updates the memory-stack. Therefore, it is not necessary to filter the entire sequence at once. This can be done individually, observation by observation, which allows it to finalize all computations of one single observation. In particular, the two update steps to assemble the normal equation system can be performed within the same loop

(cf. Alg. 1). After the filtering of the observation equation, the dyadic product of each observation is computed and added to the normal equations (update-step).

This strategy is very efficient, because all computations are done within one loop through the observations. A low-order recursive filter is an ideal choice to represent the correlations of the measurement series in order to reduce the computational effort of filtering all columns of the matrix \mathbf{A} (filter-array).

In the Sec. 4 it is mentioned that the phase spectrum $\triangleleft(\omega)$ of this filter is not linear. Therefore, the group delay (the derivative of the phase spectrum) varies for different frequencies. This individual delay is absorbed during the assembling of the normal equations $\mathbf{A}^T \mathbf{F}^T \mathbf{F} \mathbf{A}$ by the operation $\mathbf{F}^T \mathbf{F}$. In matrix formulation the causality of the filter \mathbf{F} is expressed by the lower triangular Toeplitz structure, because the matrix-vector operation $\bar{\mathbf{x}} = \mathbf{F}\mathbf{x}$ combines only elements $x(k)$, $k = 0, \dots, n$ to compute $\bar{x}(n)$. In contrast, the matrix \mathbf{F}^T reflects an anti-causal system, or in terms of filtering, a filter operation with a time-reversed sequence. The delay caused by the first filter operation is compensated by the second filter operation. This re-filtering is typically used to correct the phase shifts and to produce a zero-phase filters. Confer Koch and Schmidt (1994, pp. 59-61) or Buttkus (2000, pp. 337-338).

From the computational aspect the cascaded filtering requires no additional effort if we assemble the normal equations $\bar{\mathbf{A}}^T \bar{\mathbf{A}}$ by dyadic products of the filtered rows. But applying iterative strategies we are not interested in explicit computed normal equations. Evaluating eqns. (28) computing (26) we want to take advantage of the shape of the vector \mathbf{p} . Therefore, we avoid matrix-matrix operations and prefer the following order of computation

$$\mathbf{q} = (\mathbf{A}^T (\mathbf{F}^T (\mathbf{F}(\mathbf{A} \mathbf{p})))), \quad (27)$$

on the condition to access \mathbf{A} row-by-row. The time-reversed second filter process is problematic, because $(\mathbf{F}(\mathbf{A}(k, :) \mathbf{p}))$ produces for each row k of \mathbf{A} one filtered value $h(k)$. We certainly may collect these values and perform the second filter operation with the reversed sequence, but then we have to recompute matrix \mathbf{A} to finalize the computations at eqn. (27).

A way out of this dilemma is possible by the usage of a finite impulse response filter. The matrix representation \mathbf{F} then becomes a banded lower triangular structure and therefore the filter operation is reduced from the entire sequence to a finite number observations. In addition, we can take advantage of symmetric moving average filters (see Fig. 4), for which the direction of filtering is irrelevant. This fact is reflected by the property that all frequencies have a constant delay, which can be easily handled by a constant time shift applied to the entire sequence. This approach enables a very efficient computation of eqn. (27). Avoiding the filtering of all columns of the design matrix \mathbf{A} , because only the vectors $\mathbf{h} = \mathbf{A}\mathbf{p}$ and $\bar{\mathbf{h}} = \mathbf{F}\mathbf{h}$ have do be filtered element-by-element in contrast to filter all columns of the design matrix \mathbf{A} . The update step of the vector \mathbf{q} can be performed by

$$\mathbf{q}^{(k)} = \mathbf{q}^{(k-1)} + \mathbf{A}(k, :)^T \text{FILTER_}\bar{\mathbf{h}}(\text{FILTER_}\mathbf{h}(\mathbf{A}(k, :) \mathbf{p})), \quad (28)$$

with the initial condition $\mathbf{q}^{(0)} = \mathbf{0}$, and requires only a single pass through all observations (cf. Alg. 1, right scheme). The two filter operations act on scalars, because they are applied to sequences of scalar values.

6. Conclusions

Different strategies to handle colored noise of band-limited measurements were summarized. Gridded data sets can be treated in the time domain, as well as in the frequency domain. At first glance the regularly structured covariance matrices suggest fast collocation strategies. Applying Fourier techniques opens a very efficient way to handle the correlations in the frequency domain. But on the other hand, signal processing tools allow it to treat the spectral behavior of measurement series directly with discrete filters in the time domain. The causality of the used filters renders a way for the application of least squares adjustment by direct or iterative techniques with sequential access to the observations. Because of the sequential access also long time sequences and huge problems can be solved with these well-tailored filter techniques.

Acknowledgements

Parts of the work described in this report were done under the ESA/ESTEC contract No. 14287/00/NL/DC ('*From Eötvös to mGal+*'). I would like to thank B. Kargoll and the anonymous reviewers for many useful remarks on the manuscript.

References

- Bottino G.P., R. Barzaghi: 1993, 'Fast Collocation'. *Bulletin Géodésique*, Vol. 67, No. 2, pp. 119-126.
- Buttkus B.: 2000, 'Spectral Analysis and Filter Theory in Applied Geophysics'. Springer, Berlin.
- ESA: 1999, 'The Four Candidate Earth Explorer Core Missions - Gravity Field and Steady-State Ocean Circulation Mission'. ESA Report SP-1233(1), Granada.
- Koch K.R., M. Schmidt: 1994, 'Deterministische und stochastische Signale'. Dümmler, Bonn.
- Moritz, H.: 1980, 'Advanced Physical Geodesy'. Wichmann, Karlsruhe.
- Oppenheim A., R. Schafer: 1975, 'Digital Signal Processing'. Prentice-Hall Inc., Englewood Cliffs, New Jersey.
- Schrama E.J.O.: 1991, 'Gravity Filed Error Analysis: Applications of Global Positioning System Receivers and Gradiometers on Low Orbiting Platforms'. *Journal of Geophysical Research*, Vol. 96, No. 12, pp. 20.041-20.051.
- Schuh W.-D.: 1996, 'Tailored Numerical Solution Strategies for the Global Determination of the Earth's Gravity Field'. *Mitteilungen der geodätischen Institute der TU Graz*, Folge 81.
- Sneeuw, N.: 2000, 'A semi-analytical approach to gravity field analysis from satellite observations'. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 527, München.

II: SOLID EARTH PHYSICS

LONG WAVELENGTH SEA LEVEL AND SOLID SURFACE PERTURBATIONS DRIVEN BY POLAR ICE MASS VARIATIONS: FINGERPRINTING GREENLAND AND ANTARCTIC ICE SHEET FLUX

M. E. TAMISIEA

University of Colorado, 440 UCB, Boulder, CO 80309, USA

J. X. MITROVICA

Department of Physics, University of Toronto, 60 St. George St., Toronto, Canada, M5S 1A7

J. L. DAVIS

*Harvard Smithsonian Center for Astrophysics, 60 Garden Street MS 42, Cambridge, MA 02138,
USA*

G. A. MILNE

Department of Geological Sciences, University of Durham, Durham, DH1 3LE, U.K.

Received: 17 October 2002; Accepted in final form: 12 December 2002

Abstract. Rapid ice mass variations within the large polar ice sheets lead to distinct and highly non-uniform sea-level changes that have come to be known as ‘sea-level fingerprints’. We explore in detail the physics of these fingerprints by decomposing the total sea-level change into contributions from radial perturbations in the two bounding surfaces: the geoid (or sea surface) and the solid surface. In the case of a melting event, the sea-level fingerprint is characterized by a sea-level fall in the near-field of the ice complex and a gradually increasing sea-level rise (from 0.0 to 1.3 times the eustatic value) as one considers sites at progressively greater distances (up to $\sim 90^\circ$ or so) from the ice sheet. The far-field redistribution is largely driven by the relaxation of the sea-surface as the gravitational pull of the ablating ice sheet weakens. The near-field sea-level fall is a consequence of both this relaxation and ocean-plus-ice unloading of the solid surface. We argue that the fingerprints provide a natural explanation for geographic variations in sea-level (e.g., tide gauge, satellite) observations. Therefore, they furnish a methodology for extending traditional analyses of these observations to estimate not only the globally averaged sea-level rate but also the individual contributions to this rate (i.e., the sources).

1. Introduction

Sea-level changes are thought to be a sensitive indicator of global climate variations (Intergovernmental Panel on Climate Change, 2001), and, as a consequence, they remain a focus of active research and debate. At the most basic level, this debate involves two related questions. First, what is the mean rate of global sea-level rise over the last century (and is this rate increasing)? Second, what is the source of the sea-level change? More specifically, what are the relative contributions to the observed sea-level rise from melting of global ice reservoirs, ocean thermal expansion, variations in groundwater storage, etc?



Space Science Reviews **108:** 81–93, 2003.

© 2003 Kluwer Academic Publishers.

Estimates of global sea-level rates have been obtained primarily from analyses of tide gauge records spanning several decades or more (e.g., Peltier and Tushingham, 1989, 1991; Trupin and Wahr, 1990; Douglas, 1991, 1997; Gornitz, 1995; Mitrovica and Davis, 1995; Davis and Mitrovica, 1996). These estimates vary from $\sim 1 - 3$ mm/yr (see reviews by Gornitz, 1995; Douglas, 1997), although more recent inferences range between 1.5 – 2 mm/yr (Munk, 2002). While an acceleration of sea-level is expected from climate models (IPCC, 2001), a change in rate has not yet been robustly detected in the tide gauge record (Woodworth, 1990; Douglas, 1992). Tide gauges measure the height of the sea surface relative to the local solid surface height. Independent estimates of the former have been derived using satellite-altimetry (e.g., Nerem, 1995). These estimates, which are clearly based on a more limited and recent temporal window of observations, generally lie within the range 2 – 3 mm/yr.

While a reasonable level of agreement has emerged regarding the mean rate of global sea-level rise, there is no consensus on the primary contributor to this rise. Munk (2002) has recently reviewed various arguments that favour either polar (Antarctic, Greenland) ice melting or thermal expansion as the dominant contributor to twentieth century sea-level rise. As an example, a simple budget for oceanic heat storage weakens the thermal expansion argument. In contrast, Munk (2002) suggests that IPCC (2001) estimates of polar mass balance and a suite of observational constraints on Earth rotation that are sensitive to this mass balance (e.g., recent polar wander and length-of-day changes, as well as ancient eclipse records), weaken the argument for polar ice melting. While we do not agree with all of these arguments (for example, Munk has underestimated the considerable uncertainty in the Earth rotation signal associated with postglacial rebound), his discussion does highlight the broad issues that frame the current debate.

It is clear from this discussion that source information is generally based on the amplitude of the observed global sea-level rise. A noteworthy example of this philosophy is the recent analysis by Cabanes et al. (2001). Cabanes et al. (2001) computed the thermal expansion of the oceans using a compilation of ocean temperature profiles (Levitus et al., 2000) and argued that mean sea-level rates observed from both tide gauge and satellite data were dominated by this signal. In a series of recent articles (Mitrovica et al., 2001; Tamisiea et al., 2001) we have proposed a method for constraining source information that is based, instead, on the geographic variation of the observed global sea-level rate. The idea is straightforward. It has long been known (e.g., Woodward, 1888; Farrell and Clark, 1976) that rapid melting of polar ice complexes will lead to highly non-uniform (i.e., non-eustatic) patterns of sea-level and sea-surface change. Moreover, since these patterns arise from loading and gravitational effects, the melting of individual ice reservoirs will each yield distinct patterns, or ‘fingerprints’, and these patterns should have implications for the analysis of global sea-level observations (e.g., Clark and Primus, 1987; Nakiboglu and Lambeck, 1991; Conrad and Hager, 1997; Mitrovica et al., 2001; Plag and Jüttner, 2001; Tamisiea et al., 2001). In particular,

any geographic variation evident in these observables will reflect a combination of the fingerprints of individual ice mass variations and geographic patterns associated with other contributing processes (e.g., thermal expansion, groundwater storage variations).

In applying the methodology there are two practical considerations. First, are the fingerprints of all significant contributors to the sea-level change well constrained? Second, do the observations (e.g., tide gauge data, satellite measurements) have sufficient accuracy and geographic distribution to permit an inversion for the individual fingerprints (contributions)? As a first application (Mitrovica et al., 2001), we used a new sea-level theory (Milne et al., 1999) to compute sea-level patterns associated with melting of the Greenland and Antarctic ice sheets and a collection of mountain glaciers and small ice sheets tabulated by Meier (1984). We demonstrated that melting of the Greenland ice complex equivalent to ~ 0.6 mm/yr of sea-level rise over the last century reconciled geographic variations evident in a set of tide gauge records chosen by Douglas (1997). In this regard, the impact of the Greenland fingerprint provided an explanation for the long-standing observation that sea-level rates in Europe are anomalously low (Douglas, 1991; Shennan and Woodworth, 1992; Lambeck et al., 1998; Woodworth et al., 1999). In subsequent work (Tamisiea et al., 2001) we showed that the component of the ice-mass-flux geoid fingerprints associated with the non-uniform redistribution of meltwater will be measurable, in principle, with 5 years of GRACE satellite data.

In this paper our goal is to provide a more detailed physical theory for the origin of the sea-level fingerprints than has appeared in our previous work. For this purpose, we focus on ice mass variations on Greenland and Antarctica. Specifically, we decompose the patterns of sea-level change associated with these two fingerprints into contributions from the radial displacement of the two bounding surfaces; namely, the geoid and the solid surface. (Note that in the context of the patterns produced by ice mass variations, we use the terms geoid and sea surface interchangeably. This equivalence is not the case for thermal expansion, however, because it does not cause mass variations.) This decomposition is performed both globally and at the specific tide gauge sites selected by Douglas (1997).

2. Theoretical Background

Modern predictions of gravitationally self-consistent sea-level changes driven by ice mass variations can be traced to Farrell and Clark (1976). The ‘sea-level equation’ derived by Farrell and Clark (1976) was valid for a non-rotating Earth with elastic or viscoelastic structure that varied with depth alone and shoreline geometries that remained fixed in time. The predictions described below were generated using a revised sea-level theory that extends this earlier work to account for the influence of both shoreline migration (due to local rise and fall of sea level or

the growth and ablation of marine-based ice) and (ice plus ocean) load-induced perturbations in Earth rotation (Milne, 1998; Milne et al., 1999).

As described above, we will define sea-level $S(\theta, \phi, t)$ as the height of the sea-surface $G(\theta, \phi, t)$ (also termed absolute sea level or the geoid) relative to the position of the solid surface $R(\theta, \phi, t)$, and thus:

$$\dot{S}(\theta, \phi, t) \equiv \dot{G}(\theta, \phi, t) - \dot{R}(\theta, \phi, t), \quad (1)$$

where the superscript denotes the time rate of change, θ and ϕ are the colatitude and east-longitude of the site, and t is the time.

This paper is primarily concerned with sea-level changes in response to rapid variations in the ice budget of either the Greenland or Antarctic ice sheets. In this case, we adopt the elastic case of the new sea-level theory and, as input, we use the (radial) elastic and density structure given by the Preliminary Reference Earth Model (Dziewonski and Anderson, 1981). The sea-level predictions are generated using a pseudo-spectral algorithm (Mitrovica and Peltier, 1991) with a truncation at spherical harmonic degree and order 512.

To highlight the physics of the sea-level response, we consider simplified loading scenarios in which the ice mass variation is constant across either the Greenland or Antarctic complex. We furthermore normalize our predictions using the so-called eustatic sea-level change. As an example, consider a rate of change in Antarctic ice mass of $\Delta \dot{M}_a$ at some time t_0 . The eustatic sea-level rate associated with this mass flux is given by:

$$\epsilon_a = \frac{\Delta \dot{M}_a}{\rho_w A}, \quad (2)$$

where ρ_w is the density of water and A is the area of the ocean. If we divide both sides of equation (1) by this rate, we have for time t_0 :

$$\dot{S}_a^n(\theta, \phi, t_0) = \dot{G}_a^n(\theta, \phi, t_0) - \dot{R}_a^n(\theta, \phi, t_0), \quad (3)$$

where the subscript a denotes an Antarctic ‘scenario’ and the superscript n denotes a normalized (by the eustatic sea-level rate) quantity. Similarly, sea-level change associated with Greenland ice mass variations may be decomposed as:

$$\dot{S}_g^n(\theta, \phi, t_0) = \dot{G}_g^n(\theta, \phi, t_0) - \dot{R}_g^n(\theta, \phi, t_0), \quad (4)$$

This normalization is meaningful because the sea-level and solid surface responses are, for a fixed load geometry, a linear function of the mass flux. Consider, for example, the sea-level rate. The fields \dot{S}_g^n and \dot{S}_a^n are global ‘fingerprints’ of the (normalized) sea-level change for the case of a geographically uniform flux of Greenland and Antarctic ice mass, respectively. That is, they provide a map of the spatial variation in the sea-level response associated with each of these mass flux scenarios. The total dimensional sea-level response associated with a mass variation on these major ice complexes can be computed by summing these fingerprints, where each is weighted by the associated eustatic rate.

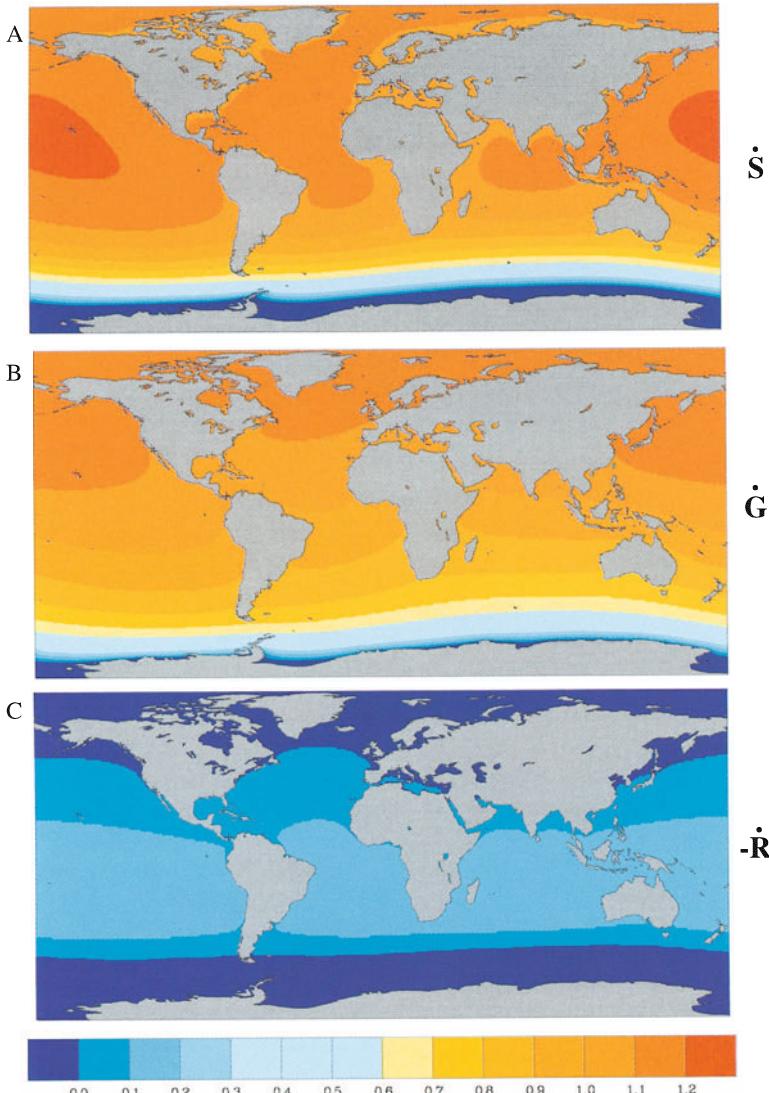


Figure 1. Normalized rates of change of (A) sea-level, $\dot{S}_a^n(\theta, \phi, t_0)$, (B) sea-surface height, $\dot{G}_a^n(\theta, \phi, t_0)$, and (C) $-\dot{R}_a^n(\theta, \phi, t_0)$, where R is the radial displacement, associated with mass flux of the Antarctic ice sheet (frame A=B+C; see equation 3). The predictions assume a uniform mass flux across the ice sheet and are normalized by the equivalent eustatic sea-level change for this mass flux. The calculations are based on the sea-level algorithm and Earth model described in the text. The small crosses on each frame indicate the location of the tide gauge sites considered in Figs. 4-5.

3. Results

Figure 1 shows a numerical prediction of the sea-level fingerprint, \dot{S}_a^n , as well as the decomposition of this field into a sea-surface (geoid) rate, \dot{G}_a^n , and radial crustal

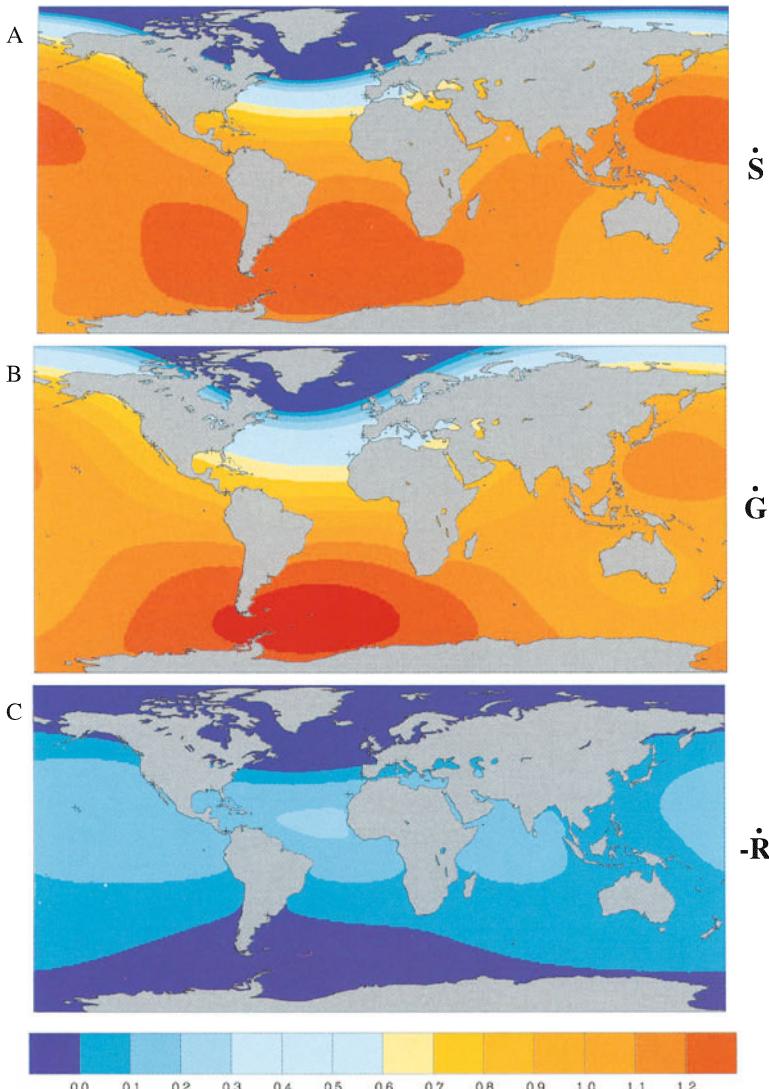


Figure 2. As in Fig. 1 except for a uniform mass flux event on the Greenland ice sheet (see equation 4).

velocity, \dot{R}_a^n , defined by equation (3). Figure 2 shows analogous fields for the case of a uniform mass flux from the Greenland ice complex (equation 4). In order to facilitate comparison between the various frames, we have chosen to plot $-\dot{R}$ rather than \dot{R} , and therefore frame (A) is the simple sum of the remaining two frames (B+C).

It is clear from Fig. 1A or 2A that growth or ablation of either the Antarctic or Greenland ice complex leads to a highly non-uniform pattern of sea-level change. As an example, let us consider the specific case of melting at a eustatic rate of

1 mm/yr from the Antarctic ice complex (i.e., the normalized value of 1 on the bottom scale of Fig. 1 refers to a dimensional sea-level rise of 1 mm/yr). In this case, the near-field of the Antarctic ice sheet experiences a sea-level fall (deep blue contour) and a gradually increasing sea-level rise as one moves from this region to the equator. In southern Chile the predicted rate is ~ 0.6 mm/yr, while the maximum predicted rise, which occurs close to Hawaii, is approximately 25% higher than the eustatic rate. The physics of this dramatic sea-level fingerprint can be explored further by considering the decomposition in Figs. 1B and C. Figure 1B indicates that the sea-surface falls in the vicinity of the ablating ice mass and rises at gradually higher rates with distance from this ice complex. Moreover, Fig. 1C shows that the melting scenario produces a large uplift in the vicinity of the ice sheets and a low amplitude (0.1-0.2 mm/yr) subsidence along a peripheral equatorial girdle. Further to the north, the subsidence tapers close to zero. (In this regard, the dark blue contour in the northern high-latitudes in Fig. 1C does not, in contrast to predictions at the southern pole, reflect a large-amplitude uplift — rather this contour represents a region of uplift that is only marginally less than zero.)

A schematic illustration of these contributions to sea-level is given in Fig. 3, which highlights (using extreme vertical exaggeration) the physics of the Antarctic sea-level fingerprint. In the top frame a large ice mass exerts a gravitational tide on the ocean and yields a sea surface which tilts gradually toward the ice complex. The ice sheet, as well as the local ‘piling-up’ of water, yields a near-field subsidence of the solid surface and a gentle uplift at the periphery. If the ice sheet melts, several inter-related processes occur. The reduction in the direct gravitational attraction associated with the ice mass leads to a migration of water from the near-field toward the far-field and hence the long-wavelength tilt of the sea-surface or geoid is reduced (see frame B). The reduction in the ice load, as well as the migration of water away from the ablating ice mass, leads to a crustal uplift in the near-field (frame B) and a gentle subsidence at the equatorial periphery.

The relative contributions to sea-level change from geoid and crustal effects are drawn roughly to scale in Fig. 3. Thus, the near-field sea-level fall predicted in Fig. 1A has comparable contributions from both the downwarping of the local geoid and the uplift of the solid surface (Figs. 1B,C and 3B). The gradual increase in the predicted sea-level rate as one moves from the near-field toward the equator (Fig. 1A) is primarily due to changes in the shape of the sea-surface. However, we note that the contribution associated with crustal deformations are not negligible in the far-field. For example, the equatorial subsidence of the solid surface is sufficient to yield a relatively flat sea-level rise from the equator to the north pole (Fig. 1A), rather than a continuously increasing rate one might expect from Fig. 1B.

The physics described above is also applicable to the Greenland mass flux fingerprint (Fig. 2A). If we consider a melting scenario in which $\epsilon_g = 1$ mm/yr, then the results in Fig. 2A indicate that sea-level will fall in regions such as Arctic Canada, Newfoundland, Norway and Scotland, and rise at rates that gradually increase from zero to 1.25 times the eustatic rate as one moves away from the

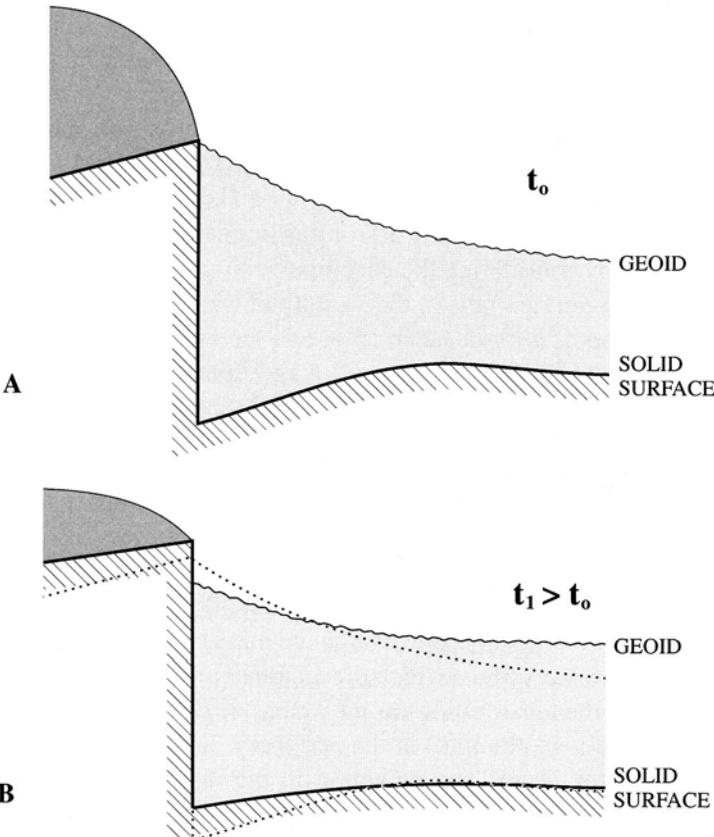


Figure 3. Schematic illustration of the origin of the sea-level fingerprints shown in Figs. 1 and 2. (A) The geoid (thin wavy line), solid surface (thick solid line) prior to the mass flux event. The light shaded region represents the ocean between these bounding surfaces. The dark shaded region at top left is a model ice sheet. (B) As in (A), except at some time t_1 after a melting event on the ice sheet. The original geoid and solid surfaces are superimposed on frame (B) as dotted lines.

Greenland ice complex. Once again, the near-field sea-level fall arises from a combination of solid surface uplift and sea-surface subsidence, while far-field patterns are dominated by sea-surface rates (compare Fig. 2A and B).

The Greenland ice mass flux introduces a sea-level contribution that is largely missing from the Antarctic fingerprint. Since the Greenland ice complex is located at a distance from the rotation pole, mass flux in the region will perturb the orientation of the rotation vector; the influence on sea-level of this perturbation contributes to the azimuthal (“quadrential”) asymmetry evident in the sea-level, sea-surface and (to a lesser extent) the radial crustal velocity fingerprints in Fig. 2.

In plotting the fingerprints we have chosen to highlight the spatial patterns within the far-field of the two polar complexes. For example, the near-field sea-level, sea-surface and crustal deformation rates reach amplitudes well beyond the

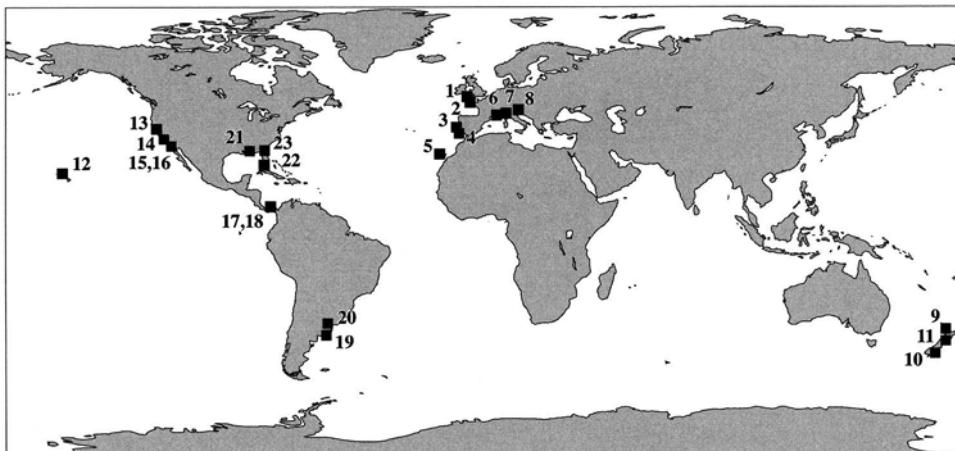


Figure 4. Location of tide gauge sites included in the analysis of *Douglas* [1997]. The name of each site may be determined by referencing the site number to the abscissa axis on Fig. 5.

scale of our colour contouring. In future work we will focus on the detailed dynamics of this near-field region (as summarized in the schematic Fig. 3). For the present purposes we are motivated by the application of the fingerprints to tide gauge analyses, and the vast majority of tide gauge sites are located in the far-field of the polar ice sheets.

To consider this issue further, we show in Fig. 4 the location of 23 tide gauge sites included in the analysis of global sea-level rise by *Douglas* (1997). (The *Douglas* analysis included a 24th site, Lyttelton, New Zealand. However, the residual tide gauge rate for this site was significantly discrepant from the other 23 and we have not included it here.) In Fig. 5 we plot the normalized sea-level rate fingerprint (solid dot), as well as contributions from the geoid rate (crosses) and radial crustal velocity (boxes) for the Antarctic (frame A) and Greenland (frame B) mass flux scenarios. Each of these sites is in the far-field of the polar ice complexes (with the exception of Newlyn and Brest relative to Greenland) and the spatial variation in the predicted sea-level rate is largely (but not solely) dominated by the fingerprint of the sea-surface variation.

Fig. 5 highlights a second, fundamentally important aspect of the fingerprint idea. Namely, sea-level fingerprints associated with melting of individual ice reservoirs have a distinct geometry (compare Figs. 1A and 2A) and thus the ability to isolate these individual fingerprints will depend on the spatial distribution of observations. Since the tide gauge sites in Fig. 4 are all at relatively large distance from the Antarctic ice sheet (with the exception, perhaps, of sites 10, 11, 19 and 20), the normalized sea-level rates predicted for an Antarctic mass flux scenario (Fig. 5A) show only moderate variations from site to site. In contrast, sites 1-8 (in Europe) and 21-23 in the eastern U.S. are sufficiently close to the Greenland ice complex that they are characterized by significantly sub-eustatic sea-level trends

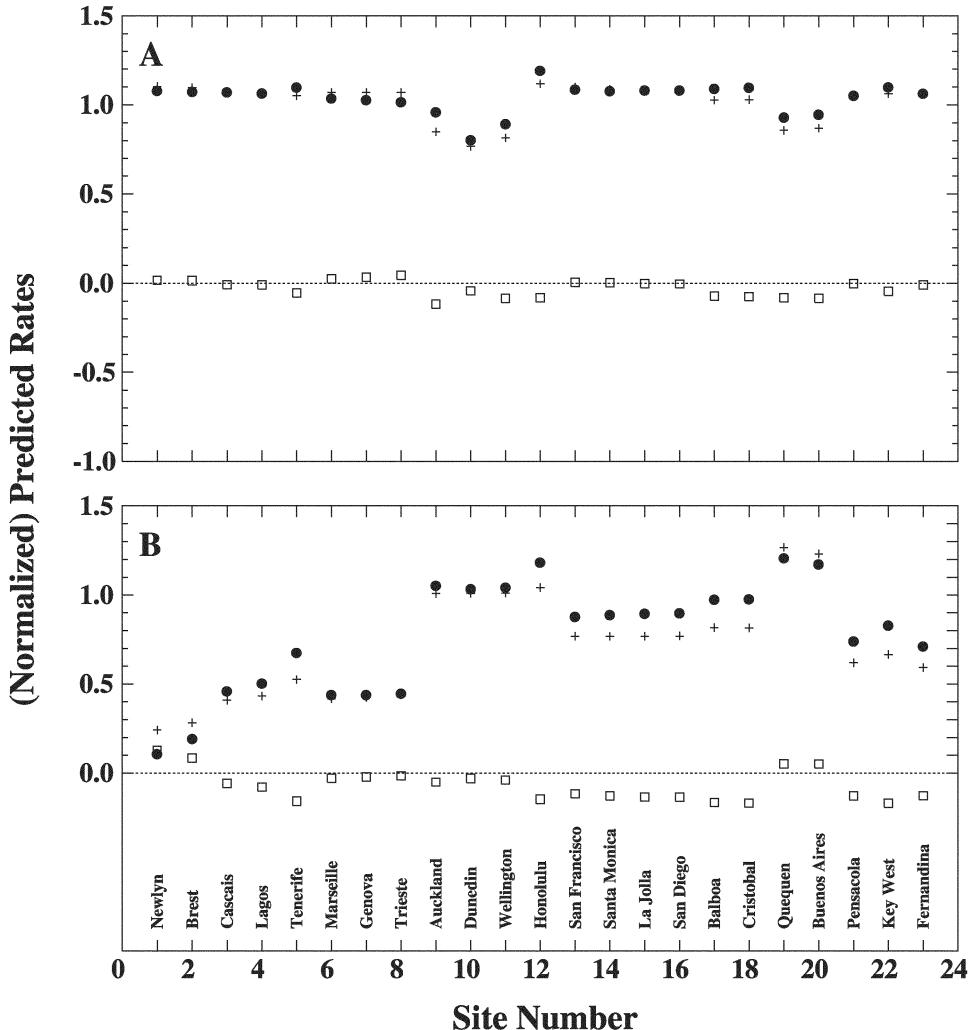


Figure 5. (A) Prediction of the normalized sea-level rate, $\dot{S}_a^n(\theta, \phi, t_0)$ (solid dots), as well as the two contributions to this rate, $\dot{G}_a^n(\theta, \phi, t_0)$ (rate of change of sea-surface or geoid height, crosses) and $\dot{R}_a^n(\theta, \phi, t_0)$ (radial crustal velocity, open squares), at each of the tide gauge sites in Fig. 4, for the case of a uniform mass flux over the Antarctic ice sheet (fingerprints for this case are shown in Fig. 1). (B) As in (A) except for a uniform mass flux event on the Greenland ice sheet (see Fig. 2).

(Fig. 5B). This raises two points. First, the data from this suite of tide gauge sites is unlikely to provide a robust estimate of Antarctic mass flux, but this data may be capable of yielding significant constraints on Greenland ice mass balance (Mitrovica et al., 2001; Tamisiea et al., 2001). Second, simple averaging of residual tide gauge rates may yield biased estimates of global sea-level rise. As an example, the mean value of the normalized sea-level rates in Fig. 5B is $\sim 0.75\%$; thus, if melting from Greenland was the only source of present-day sea-level rise, estimates of

the eustatic sea-level change generated from this set of sites would be too small by 25%.

4. Final Remarks

We have provided further physical insight into the origin of sea-level fingerprints associated with rapid mass variations on either the Antarctic or Greenland ice complex. It has been common to link these sea-level fingerprints to a simple tilting of the sea-surface. We have shown that while this tilting is dominant in the far-field of the ice complex, both sea and solid surface rates contribute to the net sea-level change in the near field (Fig. 3).

The recognition that rapid melting events give rise to dramatic spatial variations in sea-level provides a method for moving well beyond a simple estimate of the global sea-level rate. Indeed, embracing the information inherent within the geographic variation of sea-level observations allows one to estimate, in principle, both the sources of the meltwater and a global average that is unbiased by site distribution.

As a first application of these ideas, we extended a ‘traditional’ tide gauge analyses of records chosen by Douglas (1997) to consider a least-squares estimate of fingerprint weightings, and obtained an estimate of $\epsilon_g \sim 0.6 \text{ mm/yr}$ (Mitrovica et al., 2001; Tamisiea et al., 2001). While this estimate resolves an oft-cited ‘problem’ with European tide gauge rates, it must still be considered preliminary. Specifically, a robust analysis of the tide gauge trends requires a more realistic fingerprint of recent ocean thermal expansion than the simplistic eustatic pattern adopted in Mitrovica et al. (2001) and Tamisiea et al. (2001). In future work we will examine the impact on our estimate of incorporating a suite of expansion fingerprints generated from coupled ocean-atmosphere general circulation models (e.g., Gregory and Lowe, 2000).

We must emphasize, at this point, that we do not consider our approach or inferences to be in conflict with recent suggestions of a significant thermal expansion signal in sea-level observations (Cabanes et al., 2001). Cabanes et al. (2001) did not directly analyze the geographic variation in the tide gauge rates, or consider the spatial fingerprints associated with polar mass flux, just as we did not incorporate any geographic variation in the thermal expansion signal. The long-term goal of such analysis is clearly to combine well-constrained fingerprints of both ice melting reservoirs and thermal expansion together with widely distributed sea-level observations to robustly estimate the individual source contributions.

To better obtain this goal, data from near field regions would be useful. Within the near field of a given ice sheet, the amplitude (and gradient) of the signal associated with mass flux from the ice sheet is large and would dominate the contribution from distant ice sheets. Thus constraints on mass balance derived from near field

tide gauge sites will ultimately lead to improved constraints on contributions from thermal expansion.

The goal of a combined fingerprint analysis is certainly not limited to tide gauge observations of sea-level change. The fingerprints shown in Fig. 1B and 2B demonstrate that the same procedure can and should be applied to analyses of satellite-based measurements of sea-surface rates. Indeed, the great potential of the fingerprint methodology will likely be optimally realized by data collected in ongoing and future satellite missions (e.g., GRACE, CHAMP, GOCE, etc.)

Acknowledgements

We thank Bert Vermeersen for his thorough review of the original manuscript. This work was supported by NSERC, the Canadian Institute for Advanced Research - Earth Systems Evolution Program, the Royal Society of the United Kingdom, NASA grant NAG5-6068, and NSF grant EAR-9727055.

References

- Cabanes, C., A. Cazenave, and C. Le Provost: 2001, 'Sea level rise during past 40 years determined from satellite and in situ observations'. *Science*, **294**, 840–842.
- Clark, J. A. and J. A. Primus: 1987, 'Sea-level changes resulting from future retreat of ice sheets: an effect of CO₂ warming of the climate'. In: M. J. Tooley and I. Shennan (eds.): *Sea-Level Changes*. Institute of British Geographers, London, United Kingdom, pp. 356–370.
- Conrad, C., and B. H. Hager: 1997, 'Spatial variations in the rate of sea level rise caused by present-day melting of glaciers and ice sheets'. *Geophys. Res. Lett.*, **24**, 1503–1506.
- Davis, J. L., and J. X. Mitrovica: 1996, 'Glacial isostatic adjustment and the anomalous tide gauge record of eastern North America'. *Nature*, **379**, 331–333.
- Douglas, B. C.: 1991, 'Global sea level rise'. *J. Geophys. Res.*, **96**, 6981–6992.
- Douglas, B. C.: 1992, 'Global sea level acceleration'. *J. Geophys. Res.*, **97**, 12699–12706.
- Douglas, B. C.: 1997, 'Global sea level rise: A redetermination'. *Surv. Geophys.*, **18**, 279–292.
- Dziewonski, A. M., and D. L. Anderson: 1981, 'Preliminary reference Earth model (PREM)'. *Phys. Earth Planet. Inter.*, **25**, 297–356.
- Farrell, W. E., and J. T. Clark: 1976, 'On postglacial sea level'. *Geophys. J. R. astr. Soc.*, **46**, 647–667.
- Gornitz, V.: 1995, 'Sea level rise: A review of recent past and near-future trends'. *Earth Surface Processes and Landforms*, **20**, 7–20.
- Gregory, J. M., and J. A. Lowe: 2000, 'Predictions of global and regional sea-level rise using AOGCMs with and without flux adjustment'. *Geophys. Res. Lett.*, **27**, 3069–3072.
- Intergovernmental Panel on Climate Change: 2001, *Climate Change 2001: The Scientific Basis, The Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, Cambridge, UK.
- Lambeck, K., C. Smither, and P. Johnston: 1998, 'Sea-level change, glacial rebound and mantle viscosity for northern Europe'. *Geophys. J. Int.*, **134**, 102–144.
- Levitus, S., J. L. Antonov, T. P. Boyer, and C. Stephen: 2000, 'Warming of the global ocean'. *Science*, **287**, 2225–2229.
- Meier, M. F.: 1984, 'Contribution of small glaciers to global sea level'. *Science*, **226**, 1418–1421.

- Milne, G. A.: 1998, 'Refining models of the glacial isostatic adjustment process'. Ph. D. thesis, University of Toronto, Toronto.
- Milne, G. A., J. X. Mitrovica, and J. L. Davis: 1999, 'Near-field hydro-isostasy: The implementation of a revised sea-level equation'. *Geophys. J. Int.*, **139**, 464–482.
- Mitrovica, J. X., and J. L. Davis: 1995, 'Present-day post-glacial sea level change far from the Late Pleistocene ice sheets: Implications for recent analyses of tide gauge records'. *Geophys. Res. Lett.*, **22**, 2529–2532.
- Mitrovica, J. X., and W. R. Peltier: 1991, 'On postglacial geoid subsidence over the equatorial oceans'. *J. Geophys. Res.*, **96**, 20053–20071.
- Mitrovica, J. X., M. Tamisiea, J. L. Davis, and G. A. Milne: 2001, 'Polar ice mass variations and the geometry of global sea level change'. *Nature*, **409**, 1026–1029.
- Munk, W.: 2002, 'Twentieth century sea level: An enigma'. *Proc. Nat. Acad. Sci.*, **99**, 6550–6555.
- Nakiboglu, S. M., and K. Lambeck: 1991, 'Secular sea-level change'. In: R. Sabadini, K. Lambeck and E. Boschi (eds.): *Glacial Isostasy, Sea-Level and Mantle Rheology*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 237–258.
- Nerem, R. S.: 1995, 'Global mean sea level variations from TOPEX/POSEIDON altimeter data'. *Science*, **268**, 708–710.
- Peltier, W. R., and A. M. Tushingham: 1989, 'Global sea level rise and the greenhouse effect: Might they be connected?'. *Science*, **244**, 806–810.
- Peltier, W. R., and A. M. Tushingham: 1991, 'Influence of glacial isostatic adjustment on tide gauge measurements of secular sea level change'. *J. Geophys. Res.*, **96**, 6779–6796.
- Plag, H.-P., and H.-U. Jüttner: 2001, 'Inversion of global tide gauge data for present-day ice load changes'. In: *Proceedings of the Second International Symposium on Environmental Research in the Arctic and Fifth Ny-Alesund Scientific Seminar; Mem. Nat. Inst. Polar Res.* Vol. **54**, pp. 301–318.
- Shennan, I., and P. L. Woodworth: 1992, 'A comparison of late Holocene and twentieth-century sea level trends from the UK and North Sea region'. *Geophys. J. Int.*, **109**, 96–105.
- Tamisiea, M., J. X. Mitrovica, G. A. Milne, and J. L. Davis: 2001, 'Global geoid and sea level changes due to present-day ice mass fluctuations'. *J. Geophys. Res.*, **106**, 30849–30863.
- Trupin, A. S., and J. M. Wahr: 1990, 'Spectroscopic analysis of global tide gauge sea level data'. *Geophys. J. Int.*, **100**, 441–453.
- Woodward, R. S.: 1888, 'On the form and position of mean sea level'. *United States Geol. Survey Bull.*, **48**, 87–170.
- Woodworth, P. L.: 1990, 'A search for accelerations in records of European mean sea level'. *Int. J. Climatology*, **10**, 129–143.
- Woodworth, P. L., M. N. Tsimplis, R. A. Flather and I. Shennan: 1999, 'A review of the trends observed in British Isles mean sea level data measured by tide gauges'. *Geophys. J. Int.*, **136**, 651–670.

Address for Offprints: M. E. Tamisiea, University of Colorado, 440 UCB, Boulder, CO 80309, USA,
tamisiea@colorado.edu

BENEFITS FROM GOCE WITHIN SOLID EARTH GEOPHYSICS

A. M. MAROTTA (anna.marie.marotta@unimi.it)

*University of Milan, Dep. of Earth Sciences, Section Geophysics, L. Cicognara 7, I-20129 Milan,
Italy*

Received: 14 October 2002; Accepted in final form: 2 December 2002

Abstract. This contribution focuses on the results of a series of mathematical models which simulate the most important geophysical processes responsible for the deformation at crustal, lithospheric and mantle level. Post-Glacial Rebound, continent-continent collision and oceanic subduction are modelled by means of analytical and finite element models in order to emphasize the benefits in the usage of the GOCE data within Solid Earth Geophysics. Our results indicate that the new image of the Earth interior as provided by GOCE will contribute to better constrain, with respect to the present status, the physical parameters defining the outermost portion of our planet and its dynamics.

1. Introduction

A precise knowledge of the density and viscosity stratification at lithosphere and mantle level plays a major role in the modelling of many geodynamical processes in Solid Earth, such as Post-Glacial rebound, continent-continent collision and oceanic subduction. Today, information about the density structure are mainly supplied by seismic tomography, which converts seismic velocity anomalies into density anomalies. Seismic tomography allows to gain a global image of the seismic velocity anomalies of the Earth mantle. In particular, it allows to infer some global features of subduction patterns and to deduce the static component of the long wavelength geoid anomalies. However, the seismically inferred density image cannot be considered completely satisfactory. In fact, due to the lack of uniform earthquakes distribution and of a global coverage of seismic stations and to the difficulty to translate velocity into density anomalies, the density structure cannot be completely resolved at the lithosphere and upper mantle wavelengths. This limits our ability to build dynamic models that simulate the important geophysical processes that shape the surface of the Earth or induce important phenomena which interfere with the life on human beings, such as the secular components of sea level changes or the earthquakes. At the same time, features like the lateral heterogeneities within important tectonic structures, such as the sedimentary basins, are not satisfactorily resolved by present day seismic tomography.

The gradiometric data that GOCE mission is expected to provide will thus improve the knowledge of the structure of the outermost part of the planet with respect to the image obtained by seismic tomography alone. GOCE will in fact provide new direct information on the intensity of the density anomalies and their



geometries at the wavelength of the crust, lithosphere and upper mantle. A simultaneous inversion of the tomographic and gradiometric data will enhance a better resolved image of the lithosphere and upper mantle that can be used as a constraint for geodynamical models which are developed to simulate the time evolution of lithosphere and mantle processes. In fact, the lithosphere structure, both in terms of density contrasts and of anomalous geometries, as obtained by GOCE, must coincide with the final stage of the geodynamic models. Furthermore, the gravity signals due to post-glacial rebound and present day ice loss in Antarctica and Greenland, from low to high frequencies, in the range in which GOCE will play a major role, can be used to infer the depth dependence of the viscosity within the upper and the lower mantle. In the next section some geodynamic modelling approaches will be discussed in the perspective of the acquisition of the new gradiometric GOCE data.

2. Glacial Isostatic Adjustment

The first set of models that will be discussed is analytical and allows to simulate the important geodynamical process of Glacial Isostatic Adjustment (now on referred as ‘GIA’) responsible for the secular component of the change of sea level and for the time evolution of the Earth gravity field. These variations can also be affected by present-day mass instability in Antarctica and Greenland. For the GIA analysis, the multilayered, spherically stratified self-gravitating relaxation model, outlined in Vermeersen and Sabadini (1997) and based on normal mode relaxation theory, is used. The redistribution of the glacial melt water on the viscoelastic Earth is solved within the ICE-3G model, by using the spectral analysis first implemented by Mitrovica and Peltier (1991). The results shown in Fig. 1 indicate how the vertical and horizontal velocities associated with GIA are sensitive to the rheological stratification of the mantle. The three panels correspond to upper and lower mantle viscosity of $\mu_{UM} = 1.0 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel a), of $\mu_{UM} = 0.5 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel b) and of $\mu_{UM} = 0.1 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel c). All panels refer to a 80 km thick elastic lithosphere. The vertical deformation, represented by the colors, is characterized by uplifting centers where the ice sheet complexes were located and subsidence in the surroundings. The softening of the upper mantle, from panel (a) to panel (c) is responsible for the smoothing of the uplift over wider areas in proximity of the ice-sheet complexes. The horizontal velocity field is characterized by two different components. A global one, directed northward or southward with respect to the equatorial region of the Earth and due to the suction effect of the mantle material toward the deglaciated regions of the northern and southern hemispheres, and a regional one, directed radially and outwards from the center of the different deglaciation centers. The relative strengths of these components, as the intensity of this vertical motion, depend both on the viscosity ratio between the

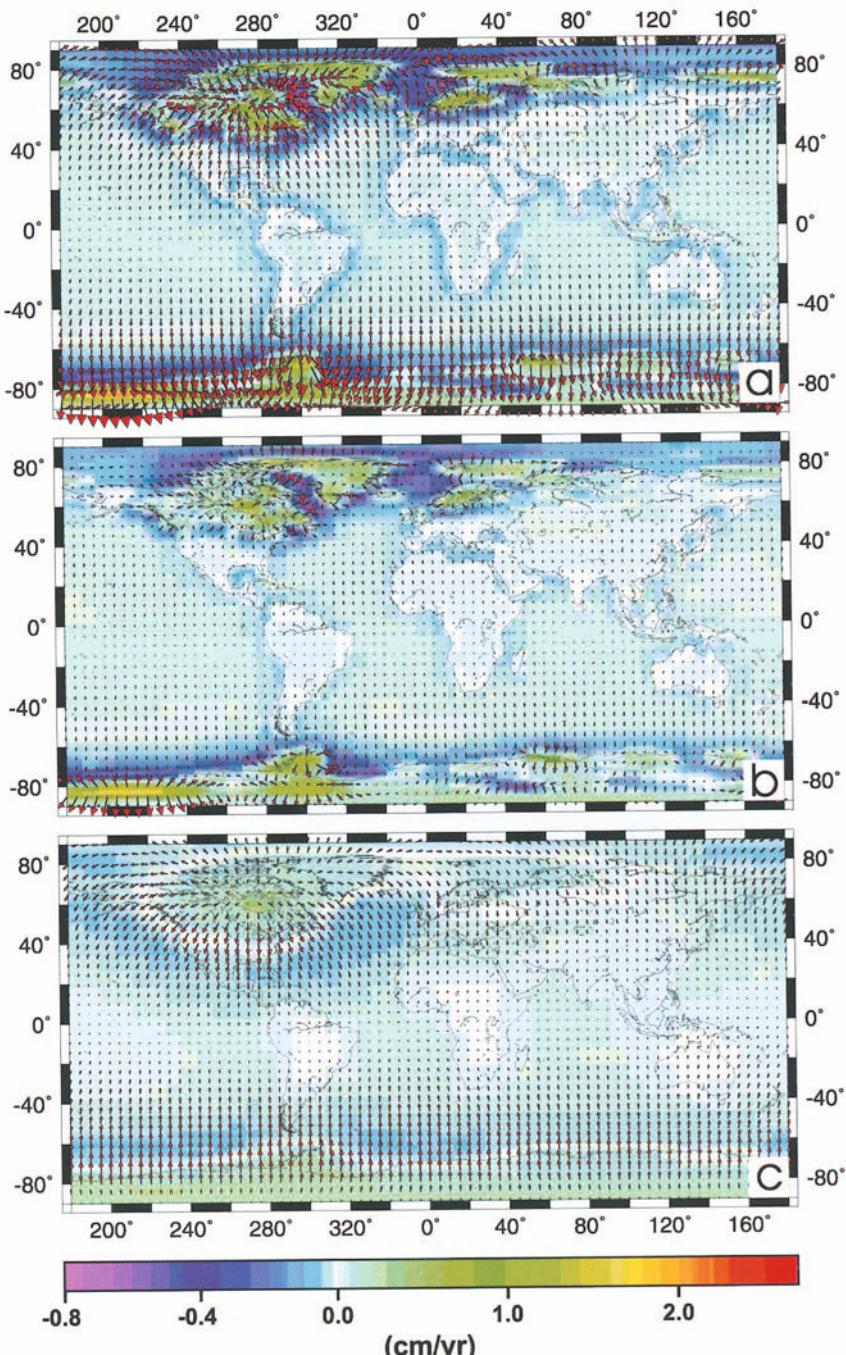


Figure 1. GIA velocity field for $\mu_{UM} = 1.0 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel a), $\mu_{UM} = 0.5 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel b) and $\mu_{UM} = 0.1 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panel c). A lithospheric thickness of 80 km is considered. Colors indicate the radial velocity while arrows indicate the horizontal one.

upper and the lower mantle and on the average global stiffness of the planet. A first general conclusion is that an increasing global stiffness of the planet induces a progressive departure from isostatic equilibrium, which is in turn responsible for higher intensity of both horizontal velocity components. Furthermore, for the same value of lower mantle viscosity, the reduction of the upper mantle viscosity with respect to lower one causes the dominance of the outward regional motion over the global equatorial flow. Thus, while for values of the upper and lower mantle viscosity ratio close to one, the north and south trending components of the horizontal velocity can exceed the local outward velocity (Fig. 1, panel a), for a low value of this ratio, the opposite occurs (Fig. 1, panel c).

3. Tectonic Deformation

The second set of models is based on a numerical scheme and allows to simulate geodynamic processes, such as oceanic subduction, intra-continental deformation and collapse of lithospheric root at a continent-continent boundary. These processes, associated with plate tectonics, involve the outermost part of the Earth and are responsible for the distribution with depth of the density anomalies that induce the gravity signals recordable by means of gradiometric space missions. These density anomalies have a crucial role in shaping the deformation and stress patterns at the Earth surface; a precise knowledge of their intensity and distribution, as it is expected from the GOCE mission, will improve the set up of these geodynamical models, in which both the density anomalies and their geometries change in time.

3.1. OCEANIC SUBDUCTION

Figure 2 portraits the predicted velocity field (arrows) overprinted on the deep temperature distribution (colors), after 10 Ma and 35 Ma of continuous oceanic subduction at a rate of 5 cm/yr, left and right coloms respectively. The modelling approach is similar to that used by Marotta and Sabadini (1995) in their study of the Tyrrhenian subduction. The flow, driven by a velocity boundary condition and density contrasts between the subducting lithosphere and the mantle, is confined in a two-dimensional cartesian geometry and the coupled set of motion and energy equations is solved for an incompressible viscous fluid. The results in Fig. 2 refer to three different rheological stratification of the upper and lower mantle. The upper and lower mantle viscosity is set to $\mu_{UM} = 1.0 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panels a and b), to $\mu_{UM} = 0.5 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panels c and d), to $\mu_{UM} = 0.5 \times 10^{21}$ Pa s and $\mu_{LM} = 1.0 \times 10^{22}$ Pa s (panels e and f). If we assume the isotherm 1500 K to define the base of the lithosphere, it can be observed that the geometry of the shallower portion of the subducted slab is mainly controlled by the kinematic regime and by the structure of the fault, somehow arbitrarily posed at the surface in order to favor the decoupling between

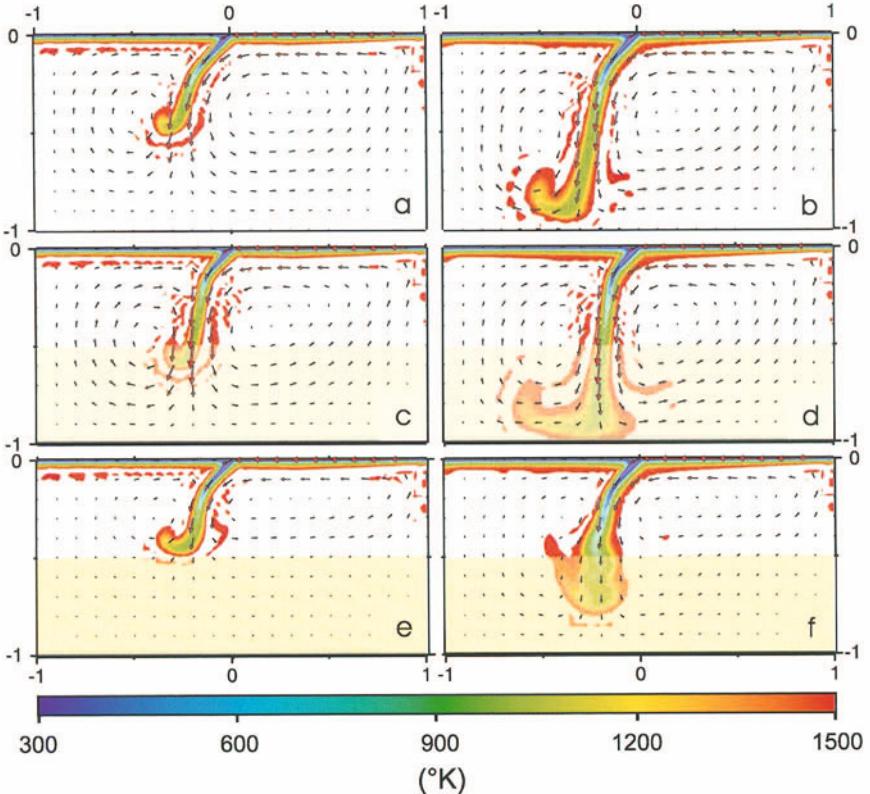


Figure 2. Velocity field (arrows) and temperature distribution (colors), after 10 Ma (panels a, c and e) and 35 Ma (panels b, d and f) of continuous oceanic subduction at a rate of 5 cm/yr. Upper and lower mantle viscosity set to: $\mu_{UM} = 1.0 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panels a and b); $\mu_{UM} = 0.5 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{21}$ Pa s (panels c and d); $\mu_{UM} = 0.5 \times 10^{21}$ Pa s, $\mu_{LM} = 1.0 \times 10^{22}$ Pa s (panels e and f).

subducting and overriding plate. In particular, the subducted slab retains a slope comparable to the dip of this fault. Varying the rheological stratification of the mantle leads, from the first to the third row, to a significant differentiation in the style of subduction at the deepest depths. For an homogeneous viscosity mantle (Fig. 2, panels a and b) after a significant time span the deepest part of the slab is almost vertical. In the deepest part of the convective model, the shape of the slab is in fact controlled by the return circulation growing in the right portion of the cell due to the velocity boundary condition at the top of the cell itself. If a softer upper mantle is introduced (Fig. 2 panels c and d), with a viscosity of $\mu_{UM} = 0.5 \times 10^{21}$ Pa s, while the lower mantle viscosity is kept at $\mu_{LM} = 1.0 \times 10^{21}$ Pa s, the global circulation is enhanced, due to the reduced global stiffness of the system, and the process becomes more rapid, with the tip of the slab that reaches the same depths earlier and with an easier penetration of the slab through the upper-lower mantle boundary (Fig. 2, panels c and d). The opposite occurs if the ratio between upper

and lower mantle viscosity is increased by one order of magnitude (Fig. 2, panels e and f); the stiffness of the system increases, leading to a smoothing in the intensity of the flow circulation, that is well restricted in the upper mantle. This is responsible for a slowdown of the process. The main diversification in the style of subduction occurs at the upper-lower mantle boundary, where the slab is now hampered to cross and, underneath this boundary a diffusive thickening of the slab occurs. The different distribution of deep density anomalies associated with different viscosity stratification would originate different gravity signals at surface. Consequently, not only GOCE will directly improve the knowledge of the deep density images, but will constrain indirectly the viscosity stratification of the mantle.

3.2. INTRA-CONTINENTAL DEFORMATION

The compilation of the World Stress Map (Zoback, 1992) shows that compressive stresses can be transmitted from plate boundaries into the interior of continental plates, leading to different types of intra-continental deformation, such as flexure of the lithosphere, folding and buckling at crustal and lithospheric level, basin inversion with reactivation of pre-existing discontinuities. Cloetingh et al. (1995, 1999) and Ziegler et al. (1995) describe tectonic settings where intra-plate compression has been observed, from North America and Africa to the Russian Platform and the Europe Alpine foreland. The late Cretaceous compressive stresses seem to have played a crucial role also in the tectonic evolution of central Europe. Several attempts have been made in the recent years to model the present—day deformation and stress fields in the European region caused by plate boundary forces and by intraplate compressive stresses (Grünthal and Stromeier, 1986; Grünthal and Stromeier, 1992; Richardson, 1992; Gölke and Coblenz, 1996; Marotta et al., 2001; Marotta and Sabadini, 2002a; Marotta and Sabadini, 2002b). Figure 3 shows the deformation field, both in terms of velocities (arrows) and vertical strain rate (colors) at the Baltica-Avalonia transition zone during the Paleocene phase of inversion, due to 35 Ma of Alpine compression at a rate of 0.4 cm/yr; these results are based on a viscous thin-sheet modelling approach similar to that used by Marotta and Sabadini (2001). In the present study, the assumption of two different structural geometries for the Baltica-Avalonia boundary (see Marotta and Sabadini, 2002b) gives remarkable different predicted horizontal and vertical deformation fields, that would mean remarkable different surficial stress field. Furthermore, since the predicted density anomalies, at crustal and lithospheric level, induce in surficial gravity signals, it is clear that GOCE data would constrain this kind of modelling approach.

3.3. MANTLE UNROOTING

In regions of plate convergence, lithospheric shortening produces crust and lithosphere thickening responsible for the gravitational destabilization of the lithospheric root, that becomes colder and thus heavier than the surrounding mantle. If the

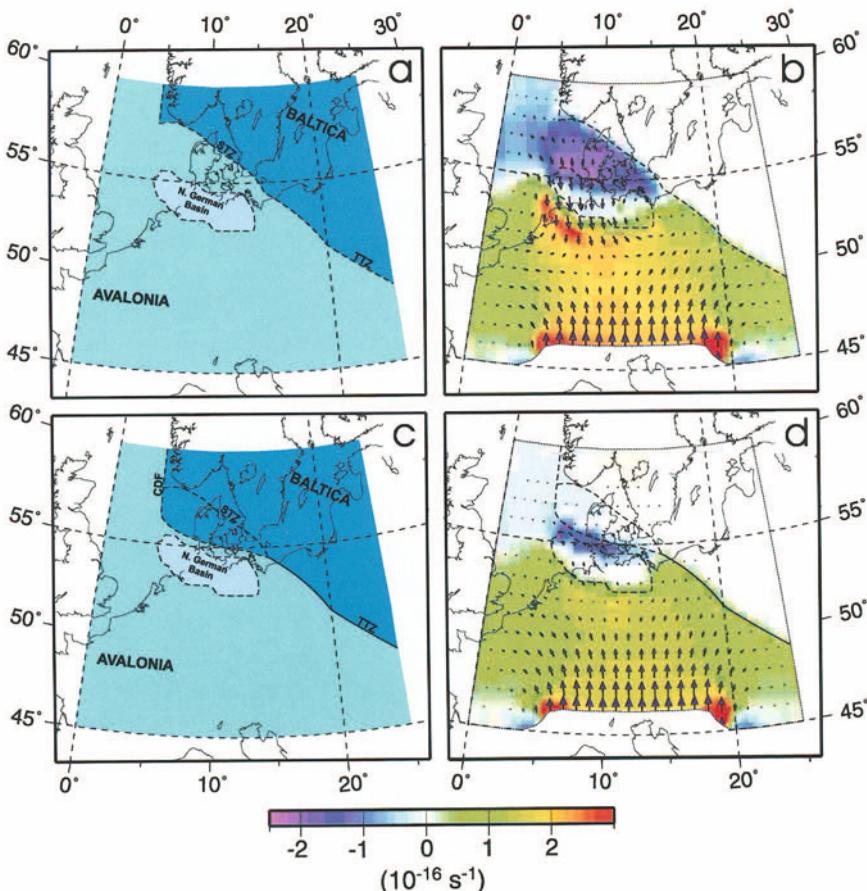


Figure 3. Sketched map of the areas between the Alpine Front and the Baltic Shield, with the main geological structures referred in the analysis (panels a and c). Predicted velocity field (arrows) and vertical strain rate (colors) after 35 Ma of Alpine compression (panels b and d). Modified after Marotta and Sabadini (2002b)

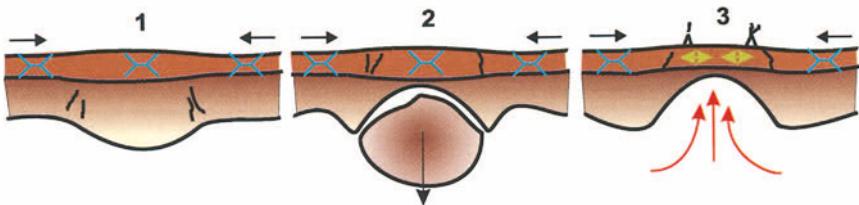


Figure 4. Schematic representation of the lithospheric unrooting process (McKenzie, 1978; Houseman et al., 1981; Fleitout and Froidevaux, 1982; Marotta et al., 1998; Marotta et al., 1999). 1. orogenic growth and initiation of gravitational instability; 2. sinking of the detached lithosphere; 3. rebound and onset of extension, coeval to still active compression. Modified after Marotta et al. (1998).

condition occurs for rock failure, the thickened lower lithospheric mantle breaks up and sinks into the sublithospheric mantle leading to a isostatic rebound responsible of a change into the surficial stress from a compressional to an extensional regime (Fig. 4). Such tectonic mechanism is often advocated to explain some controversial geological situations where compression, induced by active convergence, is observed to be coeval with extension and crustal thinning, as it seems to occur, for example, in the Alboran Sea, in the Tibetan Plateau and in the Basin and Range. Analysis of the gravitational instability of a thickened lithosphere has been developed by many authors (Houseman et al., 1981; Buck and Toksoz, 1983; Schott and Schmeling, 1998; Marotta et al., 1998; Marotta et al., 1999). All these studies enlightened the crucial role of rheology and kinematic regime on the evolution of the lithosphere-asthenosphere system when subjected to tectonic convergence. Marotta et al. (1999), in particular, demonstrated that, once the unrooting occurs, the replacement of the lower lithospheric mantle by warmer asthenospheric mantle induces a significant uplift of the surface that can precede, or be coeval with, the initiation of extension. Another important conclusion of the previously quoted study is that the intensity of the surficial uplift depends on the amount of lower lithospheric material that is involved in the unrooting process, which is in turn controlled by the rheology and the convergence velocity. The higher the convergence velocity, the bigger is the portion of the lithospheric mantle sinking into the asthenosphere (Fig. 5, panel d compared to panel b); thus, a greater dynamic topography variation occurs. Furthermore, the dynamic regime also controls the location where the uplift occurs. If the unrooting takes place during a rather slow convergence phase, the biggest increments of dynamic topography occurs in the center of the deformation region (panel a), while the opposite happens for relative high rate of convergence, with the biggest uplift at the periphery of the deformation region (panel c).

4. Conclusions

A series of geodynamic processes, related to post-glacial rebound and tectonic deformation, have been addressed to enlighten the crucial role that GOCE mission will play in improving the research within Solid Earth Geophysics. The gradiometric signals recorded during the GOCE mission will directly improve the density anomalies images, both at global and regional scales, by fulfilling the spatial gaps due to the limits of seismic tomography. Since this improved density image must coincide with the final stages predicted by the geodynamic models, GOCE will be able to better constrain this kind of models. Furthermore, it will also provide indirect information about the viscosity stratification of the mantle, which controls the time evolution of global scale processes, such as oceanic subduction and post-glacial rebound.

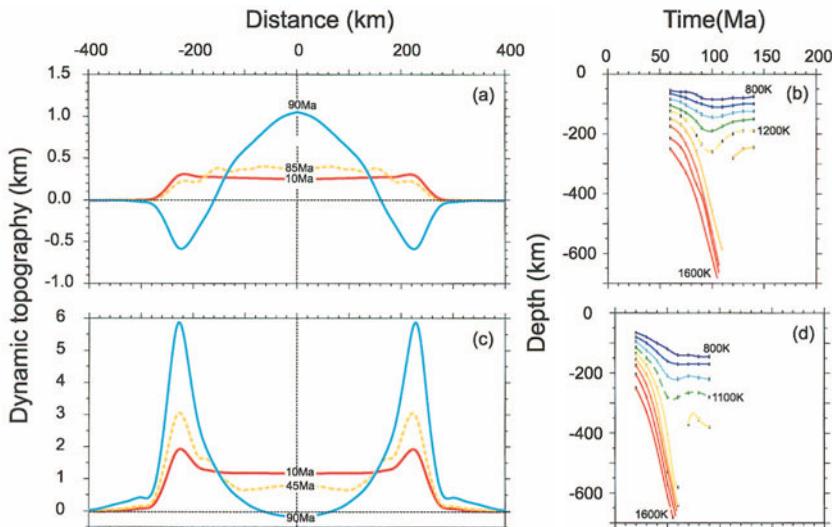


Figure 5. Dynamic topography calculated in the deformation region at different times for a rate of convergence of 0.2 cm yr^{-1} (panel a) and 1.0 cm yr^{-1} (panel c). Dashed yellow lines indicate the dynamic topography at the time when unrooting occurs. Panels (b) and (d) show the time evolution of the isotherms at the center of the thickening lithospheric root for a rate of convergence of 0.2 cm yr^{-1} and 1.0 cm yr^{-1} respectively. Dashed lines indicate the detachment levels. For a more detailed description see Marotta et al. (1999).

Acknowledgements

The author thanks Prof. Reinhard Rummel and Dr. Rudolf von Steiger for giving her the possibility to participate to the ISSI workshop on “Earth Gravity Field from Space – from Sensors to Earth Sciences”, organized in Bern, on 11–16 March 2002. The author also thanks Prof. Roberto Sabadini for the fruitful discussions during the preparation of the manuscript. Part of the results discussed in this paper was completed during the permanence of the author at the University of Bari (I), at the Institute of Earth Sciences ‘Jaume Almera’ in Barcelona (ES), at the GFZ-Potsdam (D) and at the University of Milan (I).

References

- Bonini, M., Sokoutis, D., Talbot, C. J., Boccaletti, M., and Milnes, A. G.: 1999, ‘Indenter growth in analogue models of Alpine-type deformation’, *Tectonics* **18**, 119–128.
- Buck, W. R. and Toksöz, M. N.: 1983, ‘Thermal effects of continental collisions: thickening a variable viscosity lithosphere’, *Tectonophysics* **100**, 53–69.
- Cermak, V. and Bodri, L.: 1995, ‘Three-dimensional deep temperature modelling along the European Geotraverse’, *Tectonophysics* **244**, 1–11.
- Cloetingh, S., D’Argenio, B., Catalano, R., Horváth, F., and Sassi, W.: 1995, ‘Interplay of extension and compression in basin formation’, *Tectonophysics* **252**, 1–484.
- Cloetingh, S., Burov, E., and Poliakov, A.: 1999, ‘Lithospheric folding: primary response to compression? (from central Asia to Paris basin)’, *Tectonics* **18**, 1,064–1,083.

- England, P. and McKenzie, D.: 1983, 'Correction to: A thin viscous sheet model for continental deformation', *R. Astron. Soc. Geophysic. J.* **73**, 523–532.
- Fleitout, L. and Froidevax, C.: 1982, 'Tectonic and topography for the lithosphere containing density heterogeneities', *Tectonics* **1**, 21–56.
- Gölke, M. and Coblenz, D.: 1996, 'Origin of the European regional Stress field', *Tectonophysics* **266**, 11–24.
- Grünthal, G. and Stromeyer, D.: 1986, 'Stress pattern in Central Europe and Adjacent areas', *Gerlands Beitr. Geophys.* **95**, 443–452.
- Grünthal, G. and Stromeyer, D.: 1992, 'The resent crustal stress field in Central Europe: trajectories and finite element modeling', *J. Geophys. Res.* **97**, 11,805–11,820.
- Houseman, G. A., McKenzie, D. P., and Molnar, P.: 1981, 'Convective instability of a thickened boundary layer and its relevance for the thermal evolution of continental convergent belts', *J. Geophys. Res.* **86**, 6,115–6,132.
- Marotta, A. M., Fernandez, M., and Sabadini, R.: 1998, 'Mantle unrooting in collisional settings', *Tectonophysics* **296**, 31–46.
- Marotta, A. M., Fernandez, M., and Sabadini, R.: 1999, 'The onset of extension during lithospheric shortening: a two-dimensional thermomechanical model for lithospheric unrooting', *J. Geophys. Int.* **139**, 98–114.
- Marotta, A. M., Bayer, U., Scheck, M., and Thybo, H.: 2001, 'The stress field below the NE German Basin: effects induced by the Alpine collision', *Geophys. J. Int.* **144**, F8–F12.
- Marotta, A. M., Bayer, U., Thybo, H., and Scheck, M.: 2002, 'Origin of the regional stress in the North German Basin: results from numerical modelling', *Tectonophysics*, in press.
- Marotta, A. M. and Sabadini, R.: 1995, 'The style of Tyrrhenian subduction', *Geophys. Res. Lett.* **22**, 747–750.
- Marotta, A. M. and Sabadini, R.: 2002a, 'Tectonic versus glacial isostatic deformation in Europe', *Geophys. Res. Lett.* **29**, 73-1–73-4.
- Marotta, A. M. and Sabadini, R.: 2002b, 'Numerical models of tectonic deformation at the Baltica-Avalonia transition zone during the Paleocene phase of inversion', *Tectonophysics*, in press.
- McKenzie, D. P.: 1978, 'Active tectonics of the Alpine-Himalayan belt: the Aegean Sea and surrounding regions', *Geophys. J. R. Astron. Soc.* **55**, 217–254. in press.
- Mitrovica, J. X. and Peltier, W. R.: 1991, 'On Post-glacial Geoid relaxation over the equatorial oceans', *J. Geophys. Res.* **96**, 20,053–20,071.
- Richardson, R. M.: 1992, 'Ridge forces, absolute plate motions and the intraplate stress field', *J. Geophys. Res.* **97**, 11,739–11,748.
- Schmid, S. M., Pfiffner, O. A., Froitzheim, N., Schonborn, G., and Kissling, E.: 1996, 'Geological-Geophysical transect and tectonic evolution of the Swiss-Italian Alps', *Tectonics* **15**, 1,036–1,064.
- Schott, B. and Schmeling, H.: 1998, 'Delamination and detachment of lithospheric root', *Tectonophysics* **296**, 225–247.
- Tushingham, A. M. and Peltier, W. R.: 1991, 'ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predication of postglacial relative sea level change', *J. Geophys. Res.* **96**, 4,497–4,523.
- Vermeersen, L. L. A. and Sabadini, R.: 1997, 'A new class of stratified viscoelastic models by analytical techniques', *Geophys. J. Int.* **129**, 531–570.
- Ziegler, P. A., Cloetingh, S., and Van Wees, J. D.: 1995, 'Geodynamics of intraplate compressional deformation: the Alpine foreland and other examples', *Tectonophysics* **252**, 7–61.
- Zoback, M. L.: 1992, 'First- and second-order patterns of stress in the lithosphere. The World Stress Map project', *J. Geophys. Res.* **97**, 11,703–11,728.

THE POTENTIAL OF GOCE IN CONSTRAINING THE STRUCTURE OF THE CRUST AND LITHOSPHERE FROM POST-GLACIAL REBOUND

L.L.A. VERMEERSEN (b.vermeersen@lr.tudelft.nl)
Delft Institute for Earth-Oriented Space Research, Delft University of Technology

Received: 14 May 2002; Accepted in final form: 1 August 2002

Abstract. Glacial Isostatic Adjustment (GIA) due to Pleistocene glaciation and deglaciation has left clear imprints in the present-day geoid. The solid-earth models that are commonly used in simulating these geoid anomalies usually have the upper layer (crust/lithosphere) elastic. While this is a good approximation for oceanic lithosphere, it is over-simplified for many continental crustal areas, of which some are submerged at continental margins. At many places, these continental areas have a lower crustal zone that has low viscosities. Also at the top of the mantle (asthenosphere) such zones with low viscosities can exist.

Modeling results show that, due to their shallowness and due to the laterally non-homogeneous water load, these low-viscosity layers induce discernible signatures in the high-harmonic steady-state components of the geoid. These patchlike patterns have typical length scales ranging from about 100 - 1000 km, and typical magnitudes of 1 cm - 1 m, depending on, a.o., depth and width of the low-viscosity zone, viscosity and shoreline geometry.

Complications in correlating GIA modeling results with observed geoid anomalies might arise from uncertainties in isostatic corrections (topography and non-uniform composition of crust and lithosphere) and from other non-GIA related contributions to the observed anomalies. The characteristic forms of the patterns might assist in separating the various contributions to the observed geoid anomalies. This can be illustrated for the Adriatic coast of Italy, where the best fit to the spatial sea-level curve pattern is provided by a combination of GIA and regional plate tectonics.

Keywords: crust, geoid, GOCE, lithosphere, low-viscosity solid-Earth zones, post-glacial rebound, shallow solid-Earth structure

1. Introduction

The Earth's geoid does not show much correlation with topography, at least not for low degrees and orders (less than about 30) in a spherical harmonic expansion of the Earth's gravitational potential field. For these low harmonics there is a strong correlation with dynamic processes like subduction of oceanic lithosphere, mantle convection and mantle plumes / hotspots (e.g., Crough and Jurdy, 1980; Richards and Hager, 1984; Ricard *et al.*, 1984). Also glacial isostatic adjustment (GIA) due to late-Pleistocene ice-age cycles is expected to have left its imprint in the present-day geoid, and indeed it is tempting to correlate the deep geoid low above Canada of about - 50 m with the mass deficit that causes post-glacial rebound in that region nowadays. However, it has been shown that GIA can at most contribute about half of this observed signature in that area, and that a major part of this geoid low is



linked to mantle heterogeneities (e.g., Mitrovica and Peltier, 1989; Peltier *et al.*, 1992; Simons and Hager, 1997).

Higher harmonics (degrees and order larger than about 30) are commonly linked to regional and local tectonic structures as mountains, continental plateaus, basins, etc. For these higher harmonics, geoid anomalies derived from isostatic models often give good correlations with the observed geoid. Sometimes even very simple rules can be applied: for example, for a mountain that is underlain by a low-density crustal root so that the structure is in Airy isostatic equilibrium, a rule-of-thumb of 5 m geoid anomaly for each km of topography can easily be deduced (e.g., Figures 5-20 on page 226 of Turcotte and Schubert, 1982). Still, it was known even before plate tectonics became the leading paradigm in the solid-earth sciences in the 1960's, that also for these higher harmonics there can be significant deviations of classical forms of isostasy (e.g., Chapter 10 of Heiskanen and Vening Meinesz, 1958).

A possible non-isostatic contribution to regional geoid anomalies might come from GIA for those areas that have a continental crust. Whereas oceanic lithosphere is generally strong, the continental crust and lithosphere is mostly a heterogeneous assemblage of strong layers and weak layers. Depending on the tectonic setting, stress regime and composition of the continental crust, the lower crust or crustal transition zone can have sharply reduced viscosities (e.g., Ter Voorde *et al.*, 1998). Sometimes these zones are so weak that they induce exponential-like crustal displacements after an earthquake has struck a region on time scales of years (so-called post-seismic deformation; e.g., Shen *et al.*, 1994). Many of these continental areas are submerged near continental margins. The 500 fathom contour line in bathymetry maps gives a first indication where the division is between oceanic and continental crust/lithosphere under the oceans.

The effect of such a ductile crustal zone on crustal velocities and sea-level variations induced by GIA along the U.S. east coast has been studied by Di Donato *et al.* (2000a). They found that inserting a 15 km thick ductile layer with a viscosity of 10^{18} Pa s in top elastic layers of the Earth model can induce perturbations of up to about 1 mm/yr in crustal displacements. For sea-level changes it was deduced that the incorporation of the ductile crustal zone perturbs the geographic pattern of sea-level changes along the U.S. east coast in such a way that it might account for earlier corrections in GIA models that had to be made in order to match tide gauge records.

Whereas Di Donato *et al.* (2000a) concentrated on contemporary 3-D crustal displacements and sea-level changes due to GIA for the U.S. east coast, here the long-term integrated effects of a low-viscosity ductile crustal zone on the geoid are studied. It will be shown that such intra-crustal low-viscosity zones can induce patchlike geoid anomalies due to GIA in the order of tens of centimeters to a few meters over regional distances of 100 - 1,000 km for the formerly glaciated centers and for both the near-field and far-field (with respect to the centers of the Pleistocene ice sheets) regions.

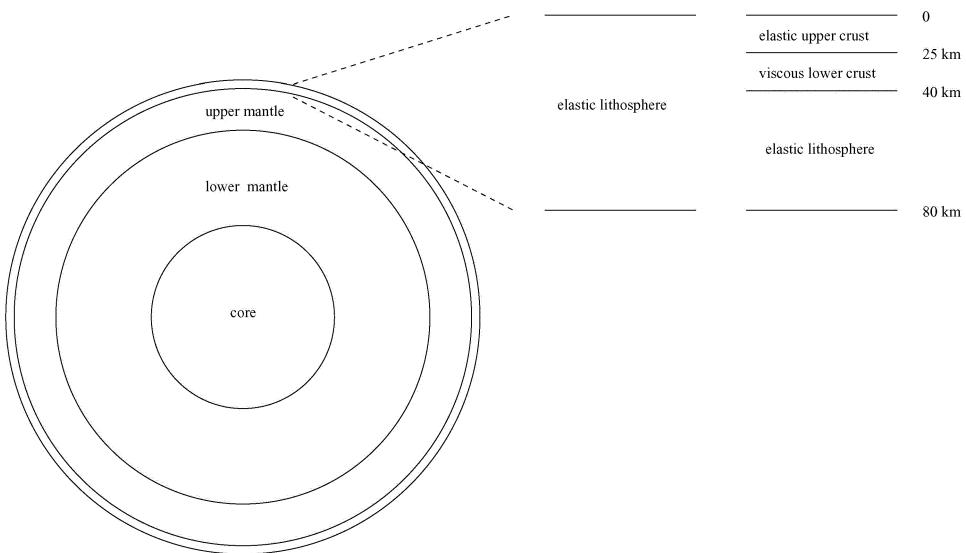


Figure 1. Earth model with (right column) and without (left column) crustal low-viscosity zone (LVZ).

2. GIA Models with Crustal Low-viscosity Zones

The Earth model used in the simulations is schematically shown in Figure 1. An inviscid core is surrounded by a viscoelastic lower and upper mantle, while the top layer consists of an 80 km thick elastic lithosphere (left column) or an elastic upper crust and lithosphere interlaced by a 10^{18} Pa s viscoelastic lower crust (right column). The elastic and constitutional parameters of the Earth model are taken from the Preliminary Reference Earth Model (PREM) of Dziewonski and Anderson (1981) and they are the same for both crust/lithosphere columns in Figure 1. Thus, the Earth model is the same as the one used in Di Donato *et al.* (2000a). The stratification and the viscosity profile for the Earth with and without the crustal low-viscosity zone is given in Table I. More details on the spherical, viscoelastic relaxation model can be found in Vermeersen and Sabadini (1997a), while the way how the geoid is determined by solving the sea level equation is described in Di Donato *et al.* (2000b).

The Pleistocene ice load model is based on the ICE-3G deglaciation history of 808 ice disks during the past 18,000 years by Tushingham and Peltier (1991). This deglaciation history is extended with seven 100,000 yr Pleistocene pre-cycles in the way described in Figure 2 of Vermeersen *et al.* (1997b).

TABLE I

Rheological structure for the Earth models without and with a low viscosity zone (LVZ)

layer	depth (km)	viscosity (Pa s) for model without LVZ	viscosity (Pa s) for model with LVZ
upper crust	0 - 25	elastic	elastic
lower crust	25 - 40	elastic	10^{18}
lithosphere	40 - 80	elastic	elastic
upper mantle	80 - 670	5×10^{20}	5×10^{20}
lower mantle	670 - 2891	5×10^{21}	5×10^{21}
core	2891 - 6371	inviscid	inviscid

3. Effects of Crustal Low-viscosity Zones on the Geoid

The effects of crustal low-viscosity zones on the present-day geoid are shown for four regions in Figure 2. Two of these regions are at the centers of Pleistocene ice sheets: the center of the Laurentide ice sheet complex in Hudson Bay, Canada (panels A and B) and the center of the Fennoscandian ice sheet complex in the Gulf of Bothnia (panels E and F). One of the regions is situated in the near-field of the Laurentide ice sheet: the east coast of the U.S. (panels C and D). And, finally, one of the regions is in the far-field of all ice sheets: the south-east Asian region (panels G and H). All these regions are at least partly underlain by continental crust/lithosphere.

The four panels in the left-hand column of Figure 2 (panels A, C, E and G) depict the present-day geoid as determined by the models described in Section 2 for the case of an elastic lithosphere. The four panels in the right-hand column (panels B, D, F and H) depict the *perturbations* of the associated panels in the left-hand column if the low-viscosity crustal zone is inserted in the solid Earth model. All units are in m.

The perturbations due to the 15 km thick low-viscosity zone are generally one to two orders of magnitude smaller than the total geoid signatures for all four regions. Whereas in the centers of the formerly glaciated areas the geoid anomalies go up to about 10 - 20 m (panels A and E), the perturbed anomalies due to the low-viscosity zone are up to about 1 m (panels B and F). For the near-field the geoid anomalies are a few m (panel C), while the perturbations due to the low-viscosity zone go up to about 10 cm (panel D). In the far-field the geoid anomalies are up to about 1 m (panel G), while the perturbations due to the low-viscosity zone go up to somewhat less than 10 cm (panel H).

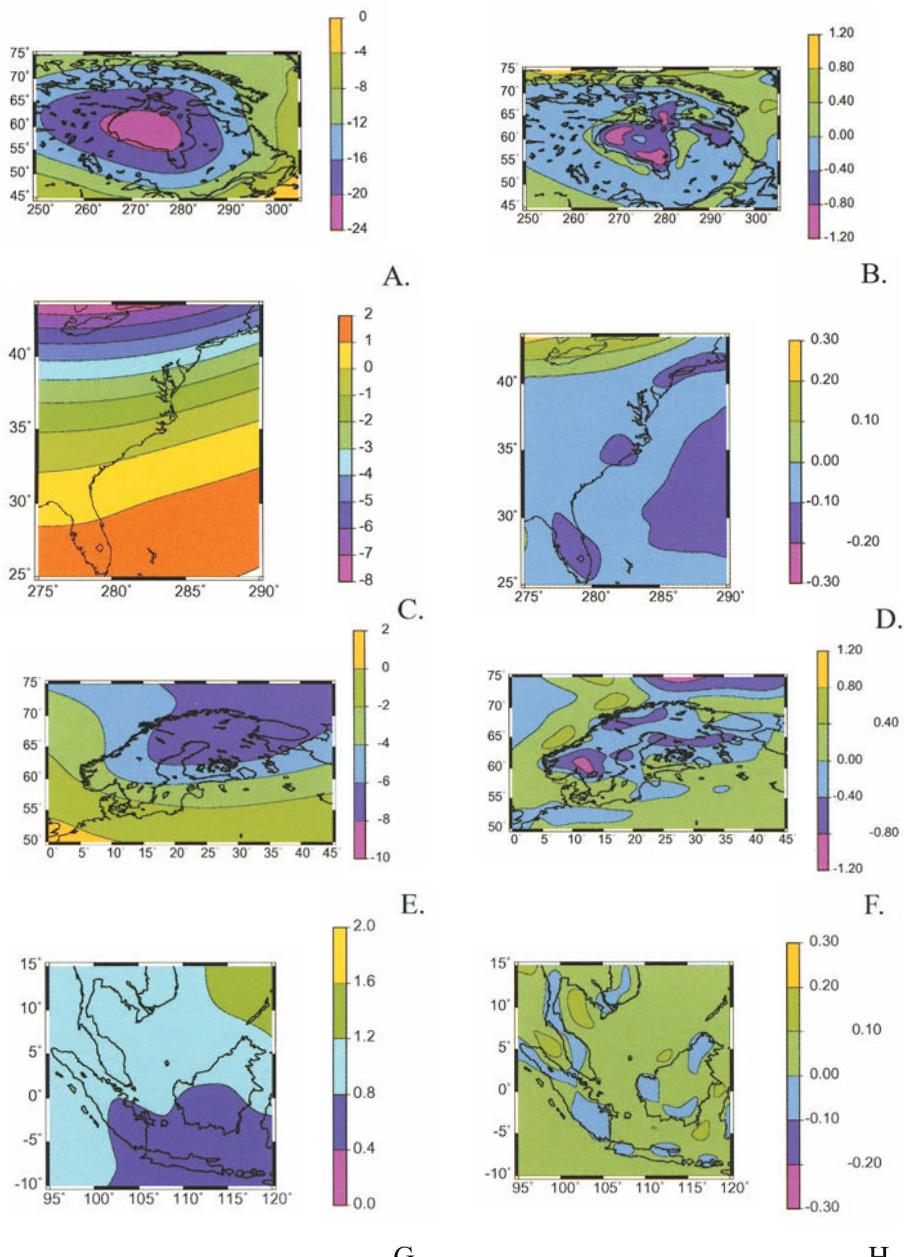


Figure 2. A) Present-day geoid (m) for Canada due to GIA with elastic crust and lithosphere. The Earth model has the properties as listed in Table I; ICE-3G has been used as Pleistocene ice load model, extended with seven growth and decay 100,000 yr pre-cycles, each cycle having a 90,000 yr ice growth phase and a 10,000 yr ice decay phase. B) Differential geoid signature due to crustal low-viscosity zone between 25 and 40 km depth. C) and D) same as A) and B), resp., for the east coast of the U.S.; E) and F) same as A) and B), resp., for Scandinavia; G) and H) same as A) and B), resp., for south-east Asia.

Concerning harmonic degrees: the panels in the left-hand column of Figure 2 show that only low harmonics are excited, which reflects the fact that the geoid anomalies are mainly caused by (upper) mantle flow. The panels in the right-hand column show that the effects of the crustal low-viscosity zone are mainly limited to high harmonics. Also this makes sense, as the low-viscosity zone is quite shallow, ranging from 25 to 40 km depth. The patchlike features that can be discerned in panels B, D, F and H have lateral extends in the range of about 100 - 1,000 km (one degree is about 110 km), although also some small contributions to higher harmonics can be seen. These patches are caused by a combination of principally two effects: the shallowness and limited radial extend of the crustal low-viscosity zone, and the ice and water load (the ocean function) distribution.

4. Detectability and Interpretation

The *Gravity field and steady-state Ocean Explorer Mission (GOCE)* of the European Space Agency, scheduled to be launched in 2006, is expected to produce an almost global uniform coverage of the geoid with a spatial resolution of about 50 - 100 km and 1 cm accuracy (Visser *et al.*, 2002). It is obvious that if these goals would be met, the geoid anomalies due to low-viscosity zones as described in Section 3 should in principle be detectable. In practice, there are many complications, both arising from uncertainties and assumptions in the GIA models and from other geophysical processes.

In Section 3 no attempt has been made to systematically study the effects of depth and width of the low-viscosity zone and of its viscosity, but it is obvious that these three variables will have their impact on form and magnitude of the geoid anomaly patches. Apart from this, there are two other major modeling assumptions that have their impact on the results: the rheology is assumed to be of the linear Maxwell type and the Earth model is assumed to have no lateral variations. The simple linear Maxwell model might not be representative for the rheology of these crustal low-viscosity layers, and perhaps the rheology is even non-linear.

The fact that the model only allows for a radial stratification and not for lateral variations might, at first sight, seem an unrealistic assumption. Certainly, in regions close to the boundary between continental and oceanic crust/lithosphere these lateral variations will have a first-order impact on the resulting anomalies. However, as the crustal low-viscosity zones are very shallow, it makes sense to assume that this first-order effect will be limited to an area of less than about 100 km from this boundary. At least, if lateral variations in continental crust thickness could be neglected. But as continental crust and lithosphere is a rather heterogeneous assemblage of different tectonic provinces, it is likely that more areas in the panels of Figure 2 will be affected by lateral variations.

Geoid anomalies are triggered by more processes than GIA, as was already indicated in Section 1. The example of the geoid low above Canada for low har-

monics (compare the - 50 m that has been observed with the modeled one in panel A of Figure 2) illustrates that mantle convection and plate tectonics can have non-negligible contributions to a region where perhaps this would not have been expected on beforehand. That such multi-component geodynamical contributions to geoid anomalies are not limited to the low harmonics, is, for instance, illustrated in the case study of historic sea-level variations of the Adriatic coast of Italy in Di Donato *et al.* (1999).

In Di Donato *et al.* (1999) it was shown that sea-level changes over the past 2,000 years, as deduced from sea-level curves derived from Roman artifacts along the east coast of Italy, can only be explained by at least two processes: GIA and plate tectonics in the Mediterranean. Along the Adriatic coast of Italy, GIA induces a trend of increasing sea-level rise from north to south, while tectonic processes in that region induce a decreasing trend in sea-level rise from north to south. Both are not representative for the observed pattern in sea-level rise; however the sum of the GIA and tectonic contributions does represent the observed inverse parabolic-like trend.

The examples of the deep geoid low above Canada and of the historic sea-level curves of the Adriatic coast of Italy show that the unique interpretation of GIA-induced geoid anomalies could be difficult. Also uncertainties in topographic corrections with classical isostasy models could hamper identification of observed geoid anomalies with the GIA-induced ones. On the other hand, the characteristic patterns in which these anomalies appear to occur (see the four panels in the right-hand column of Figure 2) might help separating and identifying the various contributions.

5. Conclusions

Apart from temporal geoid variations due to GIA and present-day continental ice-mass variations that might be observed by CHAMP and GRACE (e.g., Kaufmann, 2000; Wahr and Davis, 2002), it has been shown that low-viscosity crustal zones might induce patchlike features due to GIA in the Earth's geoid that could be observed by GOCE: magnitudes of tens of centimeters to a few meters on spatial scales of 100 - 1,000 km. Whether these patches might indeed be uniquely identified in GOCE data remains to be seen. The best chances for this will probably turn out to be for those (submerged) continental regions that are situated in relatively quiet tectonic provinces, having a smooth laterally homogeneous structure and for which high-quality seismic data are available to identify depth and width of intra-crustal low viscosity zones.

Acknowledgements

The results described in this paper are based on modeling done in close collaboration with R. Sabadini and G. di Donato of the Department of Earth Sciences of the University of Milan, Italy. The organizers of the ISSI Workshop are thanked for their cordial invitation.

References

- Crough, S.T. and Jurdy, D.M.: 1980, Subducted lithosphere, hotspots, and the geoid, *Earth Planet. Sci. Lett.*, **48**, 15–22.
- Di Donato, G., Mitrovica, J.X., Sabadini, R. and Vermeersen, L.L.A.: 2000a, The influence of a ductile crustal zone on glacial isostatic adjustment: Geodetic observables along the U.S. East Coast, *Geophys. Res. Lett.*, **27**, 3,017–3,020.
- Di Donato, G., Negredo, A.M., Sabadini, R. and Vermeersen, L.L.A.: 1999, Multiple processes causing sea level rise in the central Mediterranean, *Geophys. Res. Lett.*, **26**, 1,769–1,772.
- Di Donato, G., Vermeersen, L.L.A. and Sabadini, R.: 2000b, Sea-level changes, geoid and gravity anomalies due to Pleistocene deglaciation by means of multilayered, analytical Earth models, *Tectonophys.*, **320**, 409–418.
- Dziewonski, A.M. and Anderson, D.L.: 1981, Preliminary reference Earth model (PREM), *Phys. Earth Planet. Inter.*, **25**, 297–356.
- Heiskanen, W.A. and Vening Meinesz, F.A., *The Earth and its Gravity Field*. McGraw-Hill, New York, 1958.
- Kaufmann, G.: 2000, Ice-ocean mass balance during the Late Pleistocene glacial cycles in view of CHAMP and GRACE satellite missions, *Geophys. J. Int.*, **143**, 142–156.
- Mitrovica, J.X. and Peltier, W.R.: 1989, Pleistocene deglaciation and the global gravity field, *J. Geophys. Res.*, **94**, 13,651–13,671.
- Peltier, W.R., Forte, A.M., Mitrovica, J.X. and Dziewonski, A.M.: 1992, Earth's gravitational field: Seismic tomography resolves the enigma of the Laurentian anomaly, *Geophys. Res. Lett.*, **19**, 1,555–1,558.
- Ricard, Y., Fleitout, L. and Froidevaux, C.: 1984, Geoid heights and lithospheric stresses for a dynamic Earth, *Ann. Geophys.*, **2**, 267–286.
- Richards, M.A. and Hager, B.H.: 1984, Geoid anomalies in a dynamic earth, *J. Geophys. Res.*, **89**, 5,987–6,002.
- Shen, Z.-K., Jackson, D.D., Feng, Y., Cline, M., Kim, M., Fang, P. and Bock, Y.: 1994, Postseismic deformation following the 1992 Landers earthquake, *Seism. Soc. Am. Bull.*, **84**, 780–791.
- Simons, M. and Hager, B.H.: 1997, Localization of the gravity field and the signature of glacial rebound, *Nature*, **390**, 500–504.
- Ter Voorde, M., Van Balen, R.T., Bertotti, G. and Cloetingh, S.A.P.L.: 1998, The influence of a stratified rheology on the flexural response of the lithosphere to (un)loading by extensional faulting, *Geophys. J. Int.*, **134**, 721–735.
- Turcotte, D.L. and Schubert, D.L., *Geodynamics: Applications of Continuum Physics to Geological Problems*, John Wiley & Sons, New York, 1982.
- Tushingham, A.M. and Peltier, W.R.: 1991, ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of postglacial relative sea level change, *J. Geophys. Res.*, **96**, 4,497–4,523.
- Vermeersen, L.L.A., Fournier, A. and Sabadini, R.: 1997, Changes in rotation induced by Pleistocene ice masses with stratified analytical Earth models, *J. Geophys. Res.*, **102**, 27,689–27,702.

- Vermeersen, L.L.A. and Sabadini, R.: 1997, A new class of stratified viscoelastic models by analytical techniques, *Geophys. J. Int.*, **129**, 531–570.
- Visser, P.N.A.M., Rummel, R., Balmino, G., Sünkel, H., Johannessen, J., Aguirre, M., Woodworth, P.L., Le Provost, C., Tscherning, C.C. and Sabadini, R.: 2002, The European Earth explorer mission GOCE: Impact for the geosciences, in: *Ice Sheets, Sea Level and the Dynamic Earth*, J.X. Mitrovica and L.L.A. Vermeersen (eds), *AGU Geodynamics Series*, **29**, American Geophysical Union, Washington, 95–107.
- Wahr, J.M. and Davis, J.L.: 2002, Geodetic constraints on glacial isostatic adjustment, in: *Ice Sheets, Sea Level and the Dynamic Earth*, J.X. Mitrovica and L.L.A. Vermeersen (eds), *AGU Geodynamics Series*, **29**, American Geophysical Union, Washington, 3–32.

Address for Offprints: L.L.A. Vermeersen, Delft Institute for Earth-Oriented Space Research, Delft University of Technology, Kluyverweg 1, 2629 HS, Delft, The Netherlands, Phone: +31 15 2788272, Fax: +31 15 2785322,
email: b.vermeersen@lr.tudelft.nl

DEEP AND SHALLOW SOLID-EARTH STRUCTURES RECONSTRUCTED WITH SEQUENTIAL INTEGRATED INVERSION (SII) OF SEISMIC AND GRAVITY DATA

ROSARIA TONDI, ROBERTO DE FRANCO and GIANCARLO BIELLA

Istituto per la Dinamica dei Processi Ambientali (CNR), Via Mario Bianco, 9 Milano, Italy

(rosaria.tondi@idpa.cnr.it)

Received: 14 May 2002; Accepted in final form: 6 August 2002

Abstract. In this paper, the possibility of using simultaneously seismic and gravity data, for the reconstruction of solid-Earth structures, has been investigated through the use of an algorithm which allows joint efficient and reliable optimisation of compressional velocity and mass density parameters. We view the measured data as a realisation of a stochastic process generated by the physical parameters to be sought and we construct a “probability density function” which includes three kinds of information: information derived from gravity measurements; information derived from seismic travel time inversion and information on the physical correlation among density and velocity parameters. We show that combining data has a beneficial effect on the inversion since: it makes the problem more stable and as a consequence, providing that the quality of data is sufficiently high, enables more accurate and reliable reconstruction of the unknown parameters. In this context, we look forward the GOCE mission, which promises high spatial resolution (100–200 km) and accurate (1–2 mGals) gravity data.

We show results obtained from data sets calculated for a lateral inhomogeneous earth synthetic model and from seismic and gravity field data analysed:

- in the framework of TOMOVES (TOMOGRAphy of Mt. VESuvius) experiment, an European project aiming at reconstructing the 3-D image of Mt. Vesuvius volcano and the crust underneath, using high resolution seismic tomography techniques and other geophysical methods;
- for a profile inserted in a project aiming at reconstructing the crustal structure between Corsica and the Northern Appennines which crosses the Ligurian Sea and cuts the Ligurian Appennines W of La Spezia, extending up to Parma.

1. Introduction

The main goal of the solid-Earth sciences is to understand the past, present and future behaviour of the lithospheric component (the crust and the uppermost mantle up to depths of 80–120 km) of the earth system and the interactions with its fluid envelopes (atmosphere and hydrosphere). This is not an easy task. Earth is a very large object, very opaque, difficult to penetrate; experiments on it usually are impossible or too dangerously unpredictable to be permitted except on a very small scale. Earth scientists usually define some observable parameters whose actual values depend on the physical model parameters (like density, pressure, temperature) of interest, measure these quantities and by solving an inverse problem try to



Space Science Reviews **108**: 115–130, 2003.

© 2003 Kluwer Academic Publishers.

estimate the lithological properties of the rocks within the studied volume. For example, compressional velocities are inferred from arrival times from an earthquake or from seismic energy generated with explosive, conductivity from electromagnetic data, mass densities from gravity data. Unfortunately relationships among parameters are usually very complex, difficult to be computed and scientists have to cope with irregular and inaccurate data coverage. To address these problems it is necessary to jointly analyse the information provided by different geophysical data sets and to improve data quality and reliability. The GOCE mission promises to release high resolution, homogeneous and accurate gravity field anomalies that will give a great contribute for the analysis of the solid-Earth mass structure.

2. Methodology

2.1. THE MODEL PARAMETRIZATION

Since the starting point of the proposed method is the solution of a seismic travel times inversion problem, we have parametrized the model following the requirements of the ray tracing algorithm we have used for our purpose. The ray tracing algorithm used is that of Zelt and Smith (1992).

The velocity model parametrization is a layered, variable block (trapezoid)-size representation of a 2-D isotropic velocity earth structure. The velocities at the four corners of each trapezoid are used to interpolate a velocity field within the trapezoid so that the velocity varies linearly along its four sides. Therefore, horizontal as well as vertical velocity gradients may exist within a trapezoid. Velocity discontinuities may exist across layer boundaries, but the velocity is always continuous laterally.

As we assume that mass density and velocity contrasts occur at common boundaries, the same parametrization holds for the density model used for the computation of the density parameters adjustments. The densities at the four corners of each trapezoid are then used to compute the average density values of the block ($\rho_b = (\sum \rho_i)/4$, $i = 1, \dots, 4$) in order to estimate the gravity response.

2.2. THE INVERSION ALGORITHM

Suppose we have a starting seismic model with M total velocity points \mathbf{v}^0 and let be $\Delta\mathbf{v}(\Delta v_1, \dots, \Delta v_M)$ the velocity parameter adjustment vector calculated using a seismic travel times inversion algorithm. Here we have used a DLSQR solution (Zelt and Smith, 1992):

$$\Delta\mathbf{v} = (\mathbf{A}^T C_t^{-1} \mathbf{A} + D C_{ms}^{-1})^{-1} \mathbf{A}^T C_t^{-1} \Delta t \quad (1)$$

where \mathbf{A} is the partial derivative matrix, Δt is the travel times residual vector, C_t and C_{ms} are the estimated data and model covariance matrix and D is an overall

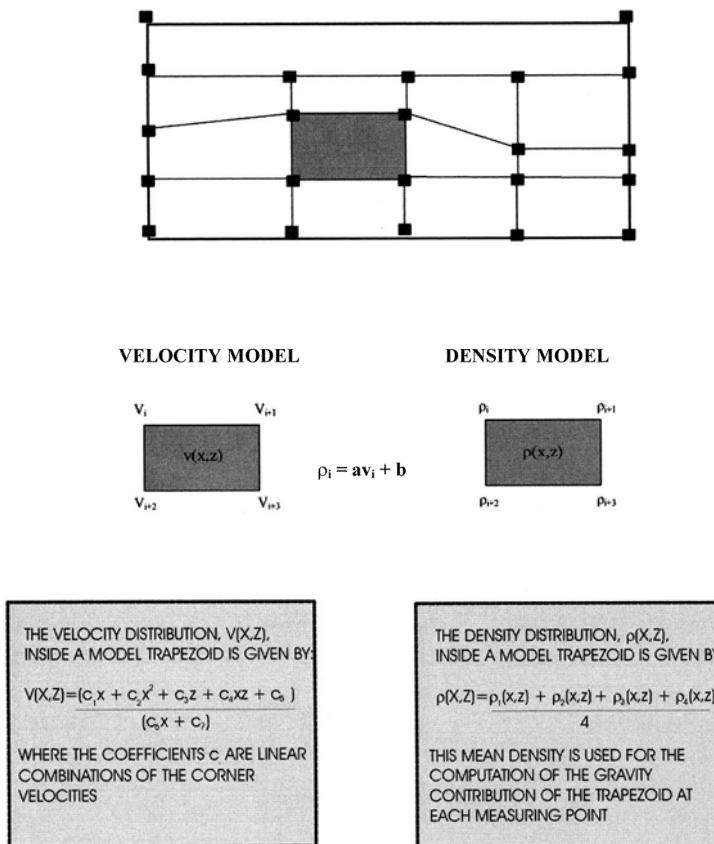


Figure 1. Velocity/density model parametrization

damping parameter. Using some linear relationship (Ludwig *et al.*, 1970; Birch, 1961):

$$\rho_m^0 = a\mathbf{v}_m^0 + b, \quad m = 1, \dots, M \quad (2)$$

with M number of total density nodes (for example for 3 adjacent blocks $M = 8$), we obtain the density model after which we can compute the gravity response in the J measuring points.

Considering that the density-velocity relationship is not linear and is partly depending on temperature, pressure and porosity, in our approach we suppose the existence of a linear relationship which may be node dependent (Tondi *et al.*, 2001). This means that we can use different relationships for different nodes.

Using the algorithm of Talwani *et al.* (1959) and considering a parametrization with B trapezoids of blocks, the gravity contribution g_j^0 , of the b -th block of

constant density at the J -th measuring point can be expressed as a product of two operators:

$$g_j^0 = \mathbf{G}_j[Z_b] \cdot \mathbf{F}[\rho_b], \quad 1 \leq b \leq B, \quad 1 \leq j \leq J \quad (3)$$

with Z_b the geometrical factor of the b -th block and r_b the density contrast of the block with respect to the background material (Telford *et al.*, 1976) or discretized in explicit form:

$$g_j^0 = 2\gamma \sum_i Z_{bij} \cdot \left(\sum_i \rho_{bi}^0 \right) / 4, \quad 1 \leq i \leq 4 \quad (4)$$

where i represents the trapezoid vertex and γ is the gravitational constant.

Comparing it with the g_j^{obs} observed values, we can then determine the gravity residuals: $\delta g_j = g_j^{\text{obs}} - g_j^0$ which account for model errors in (2) and for possible discrepancies in the starting model $[\rho_m^0]$.

Since the data are considered to be independent, random variables possessing additive, Gaussian errors with zero mean and a known covariance matrix C_x , following the regularising scheme of maximum likelihood, we can define the following “*probability density function*” (Tarantola, 1987; Tarantola and Valette, 1982) as:

$$L(\delta\mathbf{g}, \delta\mathbf{r}, \delta\mathbf{v}) \doteq \exp\{-1/2[\delta\mathbf{g} - \mathbf{G}\delta\rho]^T C_{gg}^{-1} [\delta\mathbf{g} - \mathbf{G}\delta\rho]\} \cdot \exp\{-1/2[\delta\mathbf{r} - a\delta\mathbf{v}]^T C_{mm}^{-1} [\delta\mathbf{r} - a\delta\mathbf{v}]\} \quad (5)$$

where:

$\delta\rho = \rho - \rho^0$ is the differential density parameter adjustment vector of dimension M ;

$\delta\mathbf{g} = \mathbf{g}^{\text{obs}} - \mathbf{g}^0$ is the gravity residual vector of dimension J ;

$\delta\mathbf{v} = \mathbf{v} - \mathbf{v}^0$ is the differential velocity parameter adjustment vector of dimension M ;

G is a J by M matrix containing the geometric gravitational coefficients relating each node to each gravity measurement point;

C_{gg} is the gravity data covariance matrix of dimension $J \times J$;

C_{mm} is the model covariance matrix that we assume in order to link density to velocity of dimension $M \times M$.

C_{mm} is related to the a priori linear relationship which is assumed to exist between ρ and v (2) and the error propagation is related to the uncertainty in $v(\sigma(v_m))$ and to the uncertainties in the coefficients $\alpha(\sigma(\alpha))$ and $\beta(\sigma(\beta))$:

$$\sigma(\rho_m) = v_m \sigma(\alpha) + \alpha \sigma(v_m) + \sigma(\beta), \quad 1 \leq m \leq M \quad (6)$$

Furthermore, we can consider layer information coming from the seismic/geological interpretation through the insertion of correlation between blocks which involve the

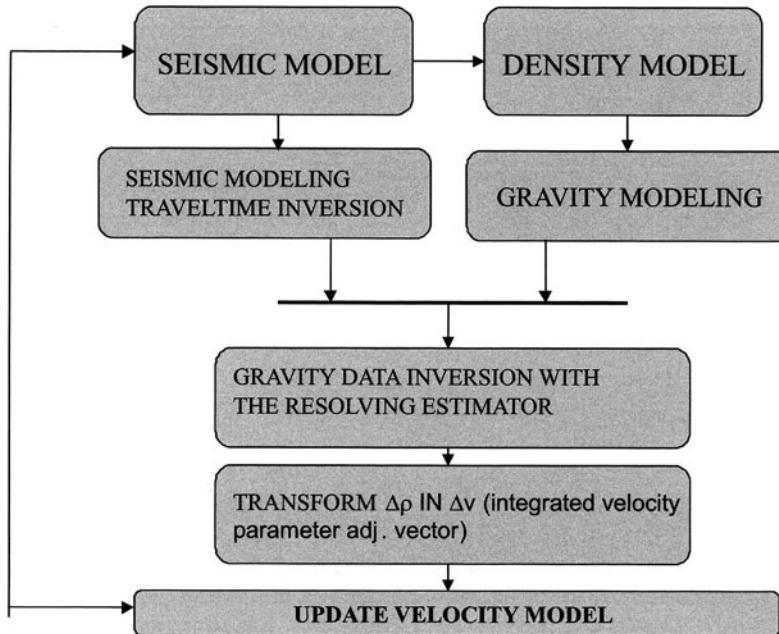


Figure 2. Flow-chart of the SII procedure

extra diagonal terms of C_{mm} . In this test of the method we have simply considered a diagonal matrix.

Once the “probability density function” is defined, we can estimate $\delta\rho$ by applying the maximum likelihood principle:

$$\max_{\delta\rho} L(\delta\mathbf{g}, \delta\rho, \delta\mathbf{v})$$

which leads to the resolving estimator:

$$\delta\rho = (\mathbf{G}^T C_{gg}^{-1} \mathbf{G} + C_{mm}^{-1})^{-1} (\mathbf{G}^T C_{gg}^{-1} \delta\mathbf{g} + a C_{mm}^{-1} \delta\mathbf{v}) \quad (7)$$

At this point, transforming $\delta\rho$ in δv through the linear differentiate relationship previously adopted: $\delta\mathbf{v} = \delta\mathbf{r}/a$, we update the velocity model.

Considering the linear relationship between ρ and v , it then follows for any block:

$$a \left(\frac{\sum_i v_i}{4} \right) + b = \frac{\sum_i (av_i + b)}{4}, \quad i = 1, \dots, 4 \quad (8)$$

and we can update the density model and estimate the new gravity residual vector. We repeat the procedure until a satisfactory fit to the observed data is achieved.

The flow-chart of the procedure is shown in Figure 2.

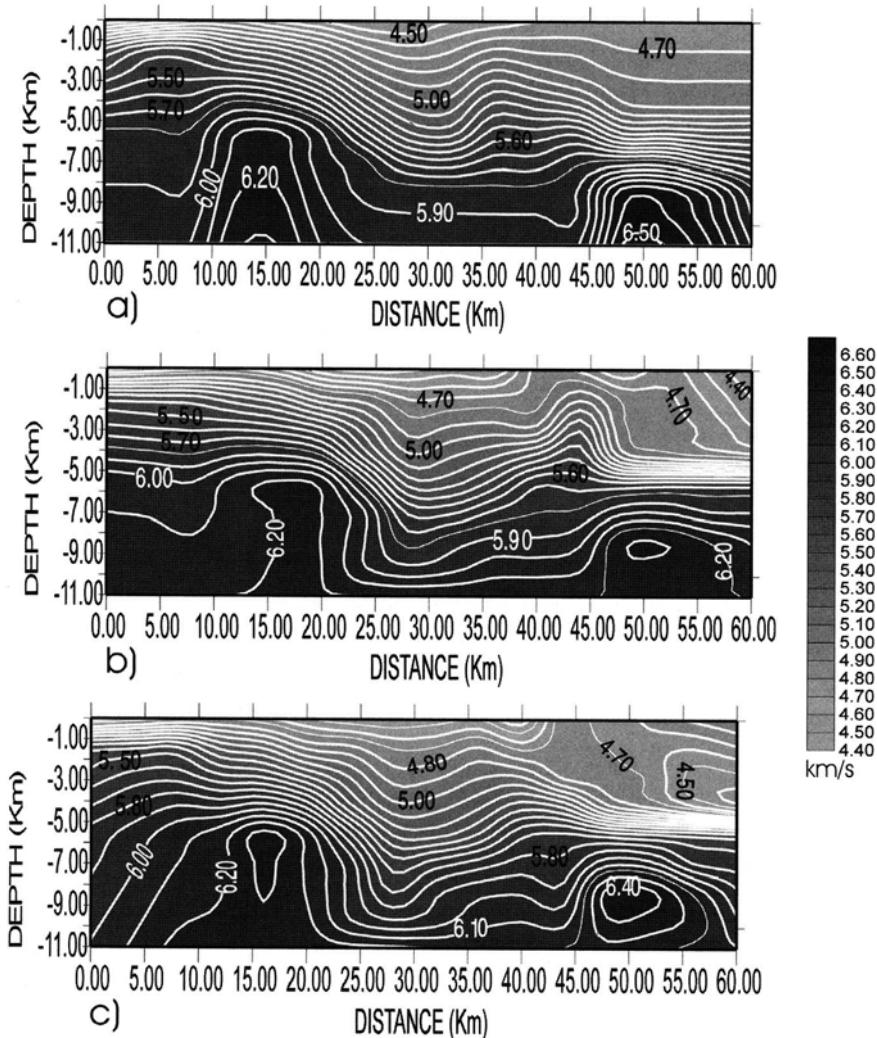


Figure 3. Comparison of synthetic model a) to model recovered after seismic travel time inversion b) and SII c)

3. Synthetic Tests

In order to test the reliability, the stability and efficiency of our sequential inversion code, we analyse the results of different inversions of data produced from a synthetic model in a way that it is possible to quantify the influence of various seismic and gravity data sets on the final solution and to characterise the differences between our inversion algorithm results and those obtained with the inversion of seismic travel times only.

The 2-D synthetic model has dimensions of 60×11 km and reproduces large and small scale earth lateral inhomogeneities (Figure 3a).

The first velocity model used to produce sets of data with different amount of information has a constant velocity for each of the four layers and the position and value of the boundary nodes (nodes on the interface between two layers) are given as “a priori” information.

Table I reports the results obtained which evidence:

- the sequential inversion solution shows a higher degree of regularity and moreover gives information in those areas not seismically resolved;
- noise added to travel times and gravity data disturb the solution, but in any case the final velocities still agree well with the true model;
- tests to verify the convergence behaviour of the method when gravity data values are lower than unknowns (underdetermined gravity information) or bigger than unknowns (overdetermined gravity information) show that it does not bring any meaningful worsening to the reproduction of the velocity field;
- the method prevents us to reach a misleading result obtained with the only optimization of travel times (we can observe from the results that even though the joint inversion increases the RMS travel time residuals, it reduces the parameter misfit).

In the next step we want to study the effect on the solution of a model parametrization (position and number of model parameters) different from the true one.

Boundary values and position have been fixed after few seismic travel times inversions.

Quantitative results are shown in Table I.

Figure 3 shows the comparison among the contoured velocity field of the true synthetic earth velocity model and the two obtained after the inversion of travel times and after the sequential inversion.

We can observe how also in this case the sequential inversion helps in dislocating masses in the right position.

4. Reconstruction of Shallow Structures: Tomoves Experiment

The SII inversion method was applied to seismic and gravity data collected in 1994 and 1996 (Gasparini *et al.*, 1998) in order to investigate the shallow structure of Mt. Vesuvius. Somma-Vesuvius is a composite volcanic complex formed by an older volcano, Mt. Somma and a younger crater, Mt. Vesuvius. The structure is located at the SE edge of the Campania Plain (Figure 4), a tectonic graben formed since Miocene times by tensional stresses acting along SW-NE and NW-SE striking fault lines (Cassano *et al.*, 1986).

Seismic signals produced by 17 on-land shots, were recorded at 220 receivers deployed along 5 profiles which have a radial distribution around Somma-Vesuvius,

TABLE I

seismic data inversion results

ITERATION	ARRIVALS	RMS (s)*	SEISMIC χ^2 **	RMS (mgal)*	PARAMETER STD (Km/s) ^{oo}	PARAMETER MISFIT (10^{-3})***
TWO SHOTS – NOISE FREE DATA						
0	75/87	0.070	49.638	6.2622	0.1536	2.91
1	79/87	0.021	4.465	3.2118	0.1293	2.45
2	72/87	0.015	2.151	3.2127	0.1392	2.64
TWO SHOTS – NOISE DATA ($\sigma_n = 0.010 - 0.025$ s)						
1	76/87	0.032	1.861	3.8123	0.1261	2.39
2	74/87	0.024	1.006	3.8154	0.1266	2.40
FOUR SHOTS – NOISE FREE DATA						
0	2110/2747	0.058	33.885	6.2622	0.1536	2.91
1	2228/2747	0.026	7.025	3.6415	0.1184	2.24
2	2365/2747	0.021	4.401	3.0277	0.1167	2.21
FIVE SHOTS – NOISE FREE DATA						
0	2567/3426	0.056	31.733	6.2622	0.1536	2.91
1	2775/3426	0.022	4.985	3.4928	0.1152	2.18
2	2763/3426	0.021	4.400	3.1571	0.1131	2.14
FIVE SHOTS: NO A PRIORI BOUNDARY NODES AND PARAMETRIZATION (LAYERS WITH CONSTANT VELOCITY)						
0	2251/3426	0.139	193.794	4.6383		
5	2312/3426	0.032	10.054	3.1440	fixed boundary nodes	
FIVE SHOTS: NO A PRIORI BOUNDARY NODES AND PARAMETRIZATION (GRADIENT MODEL)						
0	4194/4783	0.090	80.975	4.8621		
4	4444/4783	0.027	7.360	2.4807	fixed boundary nodes	

SII results

ITERATION	ARRIVALS	RMS (s)*	SEISMIC χ^2 **	RMS (mgal)*	PARAMETER STD (Km/s) ^{oo}	PARAMETER MISFIT (10^{-3})***
TWO SHOTS – NOISE FREE DATA						
0	75/87	0.070	49.638	6.2622	0.1536	2.91
1	74/87	0.042	18.068	0.1455	0.1136	2.15
2	79/87	0.037	14.123	0.1213	0.1031	1.95
TWO SHOTS – DIFFERENT BACKGROUND DENSITIES						
2.607 g/cm ³	1	70/87	0.223	502.563	0.5225	0.2408
2.64 g/cm ³	1	73/87	0.064	42.047	0.0891	0.1334
2.65 g/cm ³	1	77/87	0.047	22.633	0.0966	0.1141
2.664 g/cm ³	1	69/87	0.058	33.693	0.2113	0.1179
2.674 g/cm ³	1	80/87	0.128	166.890	0.3029	0.1359
2.707 g/cm ³	1	74/87	0.257	668.074	0.5597	0.2411
TWO SHOTS – NOISE DATA ($\sigma_n = 0.010 - 0.025$ s - $\sigma_g = 0.8$ mgal)						
1	71/87	0.055	10.008	0.8796	0.1208	2.29
2	74/87	0.048	5.270	0.9220	0.1251	4.56
TWO SHOTS – NOISE-FREE DATA: UNDERDETERMINED GRAVITY INFORMATION (130/172)						
1	74/87	0.041	16.955	0.1549	0.1142	2.16
2	77/87	0.042	17.524	0.1864	0.1132	2.14
TWO SHOTS – NOISE-FREE DATA: OVERDETERMINED GRAVITY INFORMATION (191/172)						
1	70/87	0.038	14.804	0.2114	0.1136	2.15
2	80/87	0.043	18.299	0.1893	0.1120	2.12
FOUR SHOTS – NOISE FREE DATA						
0	2110/2747	0.058	33.885	6.2622	0.1536	2.91
1	2364/2747	0.044	19.625	0.1531	0.1061	2.01
2	2194/2747	0.040	16.356	0.1377	0.0997	1.89
FIVE SHOTS – NOISE FREE DATA						
0	2567/3426	0.056	31.733	6.2622	0.1536	2.91
1	2797/3426	0.043	18.217	0.2310	0.1041	1.97
2	2763/3426	0.036	12.615	0.1626	0.0975	1.85
FIVE SHOTS: NO A PRIORI BOUNDARY NODES AND PARAMETRIZATION (LAYERS WITH CONSTANT VELOCITY)						
after fixed boundary nodes						
1	2483/3426	0.054	29.271	0.9630		
2	2521/3426	0.057	32.551	0.7347		
FIVE SHOTS: NO A PRIORI BOUNDARY NODES AND PARAMETRIZATION (GRADIENT MODEL)						
after fixed boundary nodes						
1	4368/4783	0.022	4.883	0.5551		
2	4256/4783	0.024	5.914	0.4335		

* $\left(\frac{\sum_{i=1}^n (m_i - m_{i\text{real}})^2}{N} \right)^{1/2}$ with m_i traveltimes or gravity data.

** $\sum_{i=1}^n \left(\frac{d_i - g_i(x)}{\sigma_i} \right)^2$ with d_i calculated traveltimes, $g_i(x)$ the functional relationship between the geophysical model and the predicted i -th observation, and σ_i the 'a priori' estimated uncertainty.

*** $\left(\frac{\sum_{i=1}^n (v_i - v_{i\text{real}})^2}{(\sum_{i=1}^n v_{i\text{real}}^2)} \right)^{1/2}$

oo $\left(\frac{\sum_{i=1}^n (v_i - v_{i\text{real}})^2}{N-1} \right)^{1/2}$

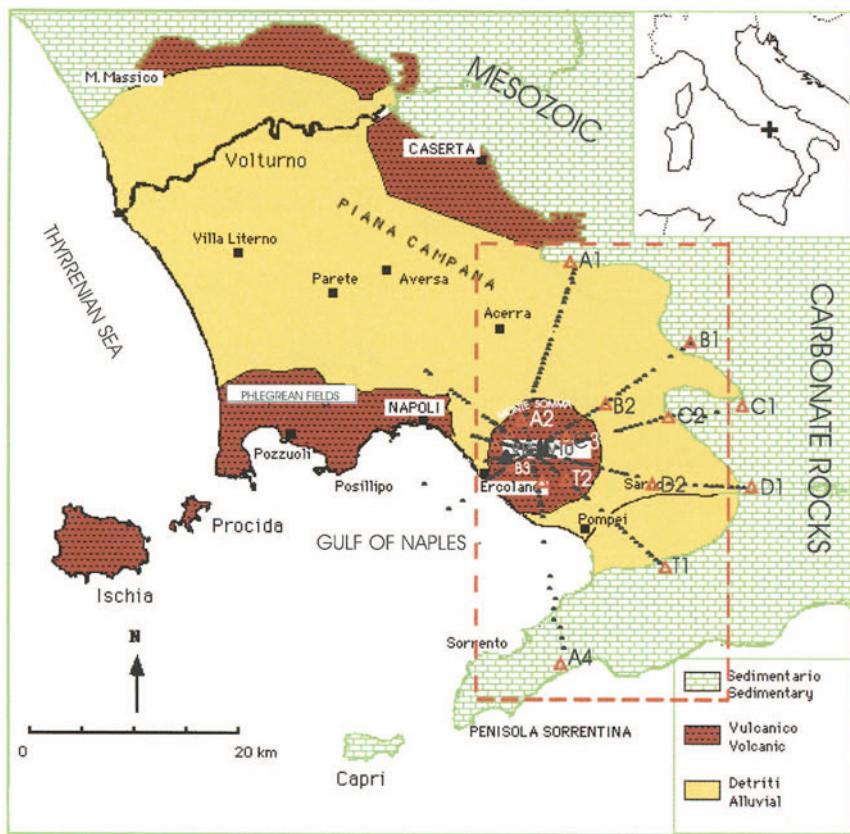


Figure 4. Geologic map of the Somma-Vesuvius region showing the five TOMOVES seismic lines. Solid triangles are shot points. The studied area by the 3-D model is outlined by a dashed rectangle.

centred near the peak of Gran Cono (Figure 4). A total amount of 550 P-wave (particle motion in the direction of seismic wave propagation) first arrivals (wave arrivals at the receiver with higher energy) through four detected geological units were used for travel time inversion. Phase P1 with a velocity ranging from 1.5 to 3.0 km/s identifies the volcanic and terrigenous unit; phase P2 with a velocity ranging from 3.0 to 5 km/s identifies the carbonates and submarine lavas; phase P3 with a velocity ranging from 5 to 6.3 km/s identifies the limestone unit; a limited number of first arrivals were observed with a velocity greater than 6.3 km/s. These are interpreted as refractions through denser limestone at depths greater than 4 km.

The gravity map of the area, which covers all the Campanian plain and the Gulf of Naples, was obtained from 17,798 measurement points. It was derived from the integration of quality checked and re-analysed previously measured on-land data (Berrino *et al.*, 1998) with new gravity data collected off-shore and along selected profiles in the Vesuvian area. As our target is the shallow structure of Mt. Vesuvius up to depths of 7 km, the gravity contribution due to deep structures is subtracted

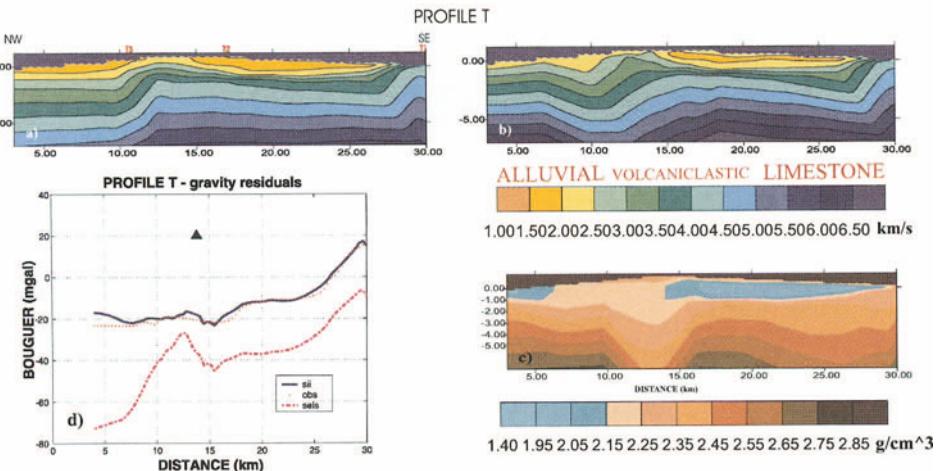


Figure 5. Velocity model along profile T recovered after seismic travelttime inversion (a) and SII (b). Velocity is contoured at 0.500 km/s intervals and related geological structures are indicated. The corresponding density model is shown in (c). In (d) is observable the reduction of gravity residuals after SII.

from the measurements of the gravity field. This contribution is calculated using collocation filtering (Barzaghi *et al.*, 1992) and the information recovered from data analysis of off-shore air gun shots performed in February 1997, in order to image the crustal discontinuities down to the Moho, below the Bay of Naples, Mt. Vesuvius and the Phlegraean Fields (Auger *et al.*, 2001; de Franco *et al.*, 2000). For the computation of Bouguer and terrain effects, a density contrast of 2.67 g/cm^3 is used.

Few travel time inversions have been performed in order to fix the position of boundary nodes. Then the subsurface of the velocity models is divided in cells of variable size and initial density models, using different velocity-density relationships depending on the velocity of the node ($\rho = 0.2587v + 1.1625$ for the first layer, $\rho = 0.2375v + 1.475$ for the second layer, $\rho = 0.17375v + 1.6625$ for the third and fourth layer) are computed. Bouguer anomaly values along the five seismic profiles are extracted by linear interpolation of values on the anomaly map. The reduction of gravity residuals (difference between observed and calculated data), together with the optimization of density models is then performed through the use of the SII procedure. Gravity residuals together with final velocity and density earth model along profile T are shown in Figure 5.

The contribution of gravity data allows to better constrain the morphology of the limestone unit to depths from 3 to 7 km, where the seismic resolution is poor or absent. In this way, we can observe, that the volcano conduit structure is inserted in this geological unit (see PROFILE T in Figure 5b and 5c between 11 and 16 km). In terms of hazard prevention we would pay attention to a relative mass deficiency observable south-eastward of Mt. Somma, consistent with the shape of gravity lows

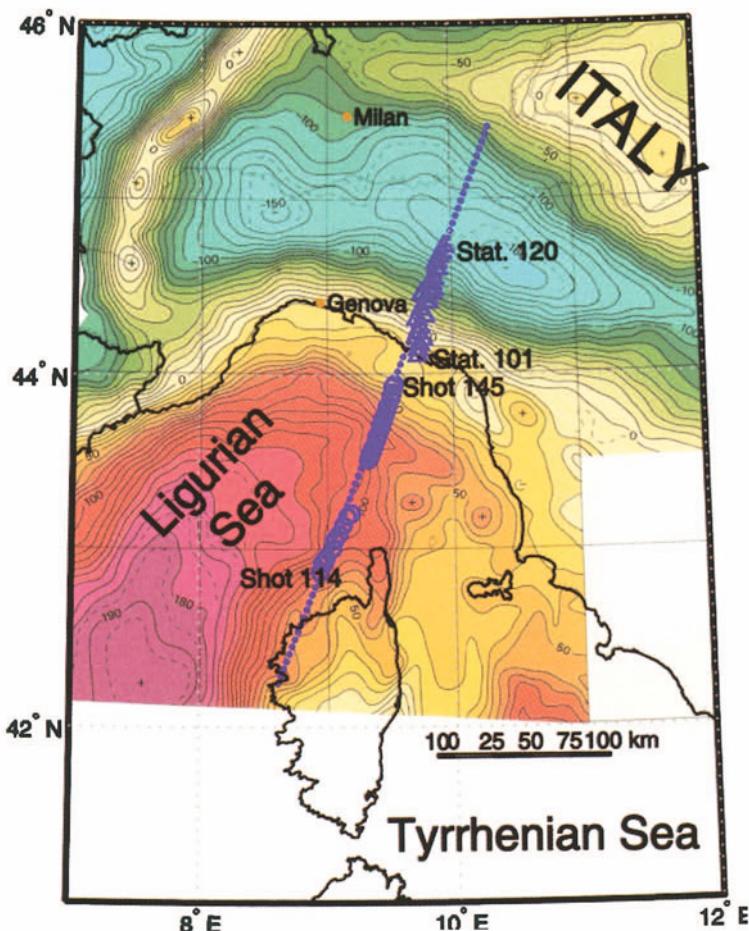


Figure 6. Bouguer anomaly map with superimposed shots and receivers.

which surround the volcano edifice and which indicate a structure through which magma may more easily flow towards the surface.

5. Reconstruction of Deep Structures: The Tuscanian-Ligurian Continental Margin

In order to verify the performance of the method in reconstructing deep solid-earth structures we have analysed data recorded in 1974 (Morelli *et al.*, 1977) along a line which begins SW of Corsica, passes the western part of the island, crosses the Ligurian Sea and cuts the Ligurian Appenines W of La Spezia extending up to Parma (Figure 6).

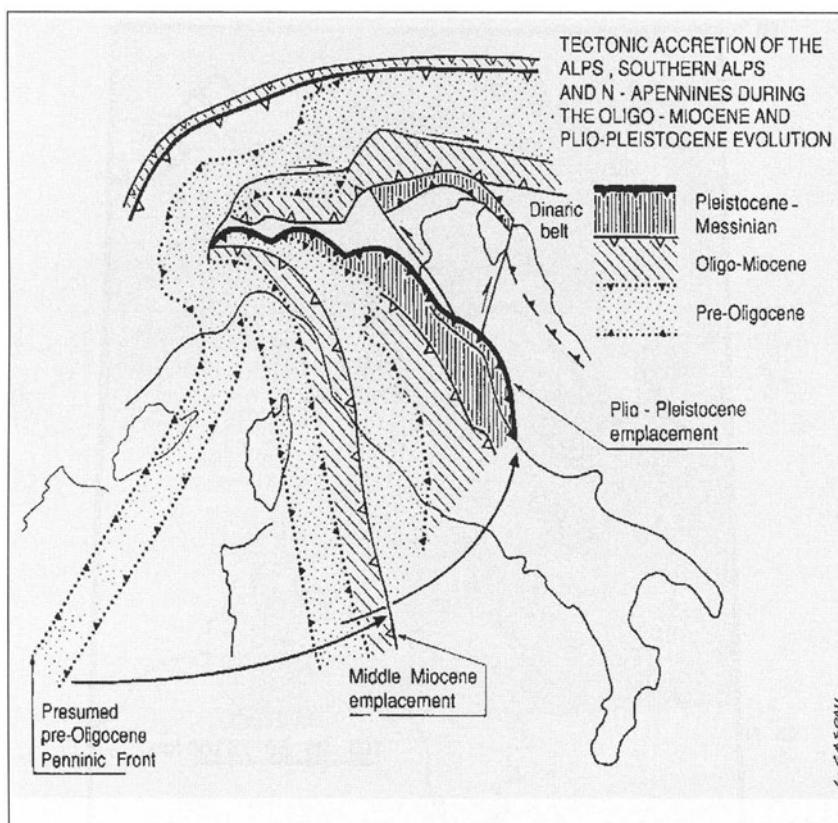


Figure 7. Evolutionary tectonic model (after Castellarin, 1992).

The Northern Apennines and their transition to the Ligurian Sea is a complex and poorly understood region with rapid changes in structural style across a short distance.

The African plate collided with Europe during the Cretaceous-Eocene, causing the closure of the Ligurian-Piedmont ocean, the formation of the Alps and the overthrusting of the Corsica Alps onto the Western Hercynian Corsica. This compressional episode was followed by a large extensional regime that occurred in the European Plate with the formation of oceanic crust in the Balearic and Ligurian seas and coeval counterclockwise rotation of the Sardo-Corso Block (22–19 Myr). The subduction of the oceanic crust ceased and at Sisco (Eastern Corsica) lamproitic rocks were emplaced around 15 Myr in an extensional tectonic regime, while compression migrated toward East-Northeast, in the inner Apenninic chain and then in the present day front of deformation under the Po plain (Doglioni *et al.*, 1998; Castellarin, 1992; Amato *et al.*, 1998; Bartole *et al.*, 1991; Makris *et al.*, 1999; Mauffret *et al.*, 1999; Tomaselli *et al.*, 1992) (Figure 7).

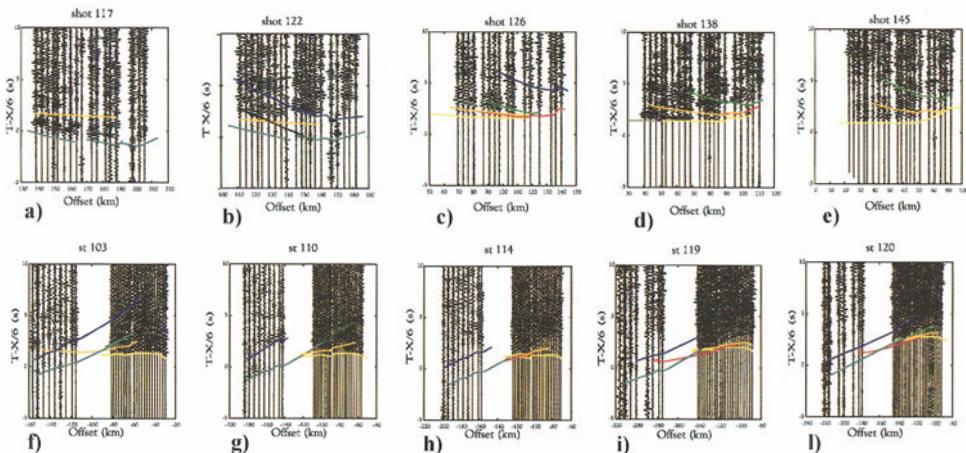


Figure 8. Selected examples of common shots (a...e) and common receiver (f...l) seismic sections. Picked phases are superimposed on recorded data. Pg arrivals correspond to yellow lines, Pi arrivals to red lines, Pn arrivals to dark green lines, PiP arrivals to orange lines, PmP arrivals to blue lines.

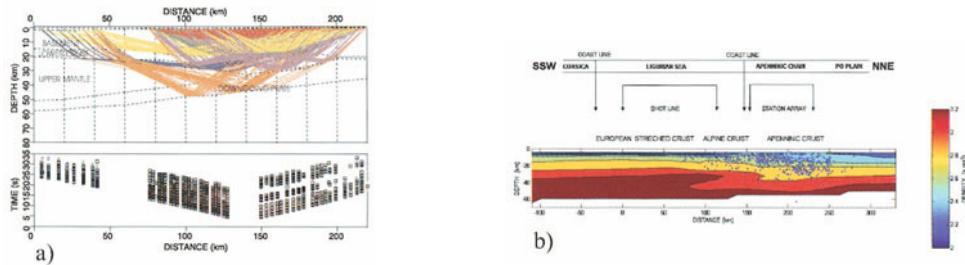


Figure 9. Ray coverage plus observed (+) and calculated (squares) traveltimes of the final velocity model as recovered after SII (a). Corresponding density model with ISC earthquakes (1970–2002) superimposed.

Seismic energy used for the experiment was generated by 32 sea shots at sites with an average spacing varying from 1.5 to 5 km. Signals were recorded by 20 on land stations deployed on a profile extending from La Spezia (station 101) to Parma (station 120).

Data, which were recently digitized and processed, were gathered in common shots (signals from one shot) and common receiver (signals from few shots received from one receiver) seismic sections (Figure 8). A total amount of 964 P-wave arrivals corresponding to Pg phase (refracted energy through the carbonatic/cristalline basement), Pi phase (refracted energy through the lower crust), Pn phase (refracted energy through the upper mantle), PiP phase (intracrustal reflections), PmP phase (wide angle reflections off the crustal/mantle boundary of the downgoing European plate) were used for travel time inversion (Figure 9a). Seismic data provide thickness and velocity control for the lower crust of the central-southwestern part of the model. The velocity crustal structure of the northeastern

part of the model is mainly constrained by the results of a DSS (Deep Seismic Sounding) experiment carried out in 1983 and 1985 by the EGT (European Geotraverse) (Egger, 1990). The southwestern part of the model is constrained by the results of Egger *et al.* (1988) on EGT data. Upper mantle velocity is constrained by reversed Pn observations beyond 140 km offset.

Gravity data are extracted from a EGT compilation produced in 1990 and distributed on a grid of 3×3 km (Blundell *et al.*, 1992). Bouguer and terrain effects have been computed with a density contrast of 2.67 g/cm^3 . The Bouguer anomaly map of the region is shown in Figure 6 and the values range from positive values of about 100 mGal in the Ligurian Sea to minimum values of -150 mGal in the core of the Apenninic chain. The interpretation is based on 182 measurement points along the profile. The gravimetric contribution points out the thickening of the crust in the central part of the earth profile to about 30–35 km. (Figure 9). This crustal thickening should be related to Africa-Europe collision during cretaceous-eocene times to which followed the closure of the Ligure-Piedmont sea. Indication of this process is given also by earthquake locations in the subsurface (Figure 9b).

Two other features evidenced by the modelling are:

- between Corsica and the western continental margin a stretched european foreland, 18–22 km thick, interpreted as the result of the large extensional regime occurred 22–19 Myr;
- the northeastward crustal structure, characterized by a crustal thickening up to 40 km, related to the present day Apenninic front of deformation under the Po plain.

6. Conclusions

Inversion of geophysical data is non-unique. However SII recovers velocity and density models which are consistent both with seismic and gravity data. And in particular the procedure:

- controls the gravity data inversion without adding any physical-mathematical constraint;
- contributes in singling out the not sampled areas by seismic rays;
- restricts the dependence of the inversion on the model parametrization;
- reduces the possibility to reach a misleading result obtained with the only inversion of seismic travel times or gravity data;
- can be applied to any seismic estimator even though it is highly dependent on its reliability;
- handles in a stochastic way the debated problem of the usefulness of considering a linear relationship between seismic velocities and densities, through the use of a node dependent linear relationship and the use of a model covariance matrix.

In this way we evidence how gravity data information can be important for a reliable reconstruction of solid-earth structures.

References

- Amato, A., Azzara, R., Basili, A., Chiarabba, C., Ciaccio, M. G., Cimini, G. B., Di Bona, M., Frepoli, A., Hunstad, I., Lucente, F. P., Margheriti, L., Mariucci, M. T., Montone, P., Nostro C., and Selvaggi, G.: 1998, 'Geodynamic evolution of the Northern Apennines from recent seismological studies', *Mem. Soc. Geol. It.* **LII**, 337–343.
- Auger, E., Gasparini, P., Virieux, J., Zollo, A.: 2001, 'Seismic evidence of an Extended Magmatic Sill Under Mt. Vesuvius', *Science* **294**, 1510–1512.
- Bartole, R., Torelli, L., Mattei, G., Peis, D., and Brancolini, G. : 1991, 'Assetto stratigrafico-strutturale del Tirreno Settentrionale: stato dell'arte', *Studi Geologici Camerti* 1991/1, 115–140.
- Barzaghi, R., Gandino, A., Sansò, F., Zenucchini, C.: 1992, 'The collocation approach to the inversion of gravity data', *Geoph. prospecting* **40**, 451–492.
- Berrino, G., Corrado, G., Riccardi, U.: 1998, 'Sea gravity data in the Gulf of Naples: a contribution to delineating the structural pattern of the Vesuvian area', *J. Volc. geotherm. Res.* **82**, 139–150.
- Birch, F.: 1961, 'Composition of the Earth's mantle', *Geophys. J. Royal Astr. Soc.* **4**, 295–311.
- Blundell, D., Freeman, R., Müller, S.: 1992, 'A continent revealed: The European Geotraverse', Cambridge University Press.
- Cassano, E., Anelli, L., Fichera, R., and Cappelli, V. : 1986, 'Pianura padana-Interpretazione integrata di dati geofisici e geologici', *73 Congr. Soc. Geolog. It.* AGIP (ed.), 1–27.
- Castellarin, A.: 1992, 'Introduzione alla progettazione del Profilo CROP 1', *Studi Geologici Camerti* 1992/2, 9–15.
- de Franco, R., Caielli, G., Tondi, R., Barzaghi, R., Biella ,G., (b) and TOMOVES group.: 2000, 'Deep structures in the Vesuvian area (Italy)', XXV General Assembly EGS, Nice, 25–29 April 2000.
- Doglioni, C., Mongelli, F., and Pialli, G.: 1998, 'Boudinage of the Alpine belt in the Apenninic back-arc', *Mem. Soc. Geol. It.* **LII**, 457–468.
- Egger, A.: 1990, 'A comprehensive compilation of seismic refraction data along the Southern segment of the European Geotraverse from the Northern Apennines to the Sardinia Channel (1979–1985)', Open file report, ETH Höngger, Zürich.
- Egger, A., Demartin, M., Ansorge, J., Banda, E., Maistrello, M.: 1988, 'The gross structures under Corsica and Sardinia', *Tectonophysics* **150**, 363–389.
- Gasparini, P., and Tomoves Group: 1998, 'Looking Inside Mt. Vesuvius', *EOS* **79**, 19,229–232.
- Ludwig, J. W., Nafe, J. E., and Drake, C. L.: 1970, 'Seismic refraction', in A. E. Maxwell, (ed.), *The Sea* **4**, Wiley, New York, pp. 53–84.
- Makris, J., Egloff, F., Nicolich, R., and Rihm, R.: 1999, 'Crustal structure from the Ligurian Sea to the Northrn Apennines-a wide angle seismic transect', *Tectonophysics* **301**, 305–319.
- Mauffret, A., Contrucci, I., and Brunet, C.: 1999, 'Structural evolution of the Northern Tyrrhenian Sea from new seismic data', *Marine and Petroleum Geology* **16**, 381–407.
- Morelli, C., Giese, P., Carrozzo, M. T., Colombi, B., Guerra, I., Hirn, A., Letz, H., Nicolich, R., Prodehl, C., Reichert, C., Röwer, P., Sapin, M., Scarascia, S., Wigger, P.: 1977, 'Crustal and upper mantle structure of the Northern Appennines, The Ligurian Sea and Corsica, derived from seismic and gravimetric data', *Bollettino di Geofisica Teorica e Appl.*, **75–76**, 199–260.
- Talwani, M., Worzel, J. L., and Landisman, M.: 1959, 'Rapid gravity computations for two-dimensional bodies with application to the Mendocino submarine fracture zone', *Journal of Geophysical Research* **64**, 49–59.
- Tarantola, A.: 1987, *Inverse Problem Theory*, Elsevier.

- Tarantola, A., and Valette B.: 1982, 'Inverse problems=quest for information, *Journal of Geophysics* **50**, 159–170.
- Telford, W. M., Geldart, L. P., Sheriff, R. E., and Keys, D. A.: 1976, *Applied Geophysics*, Cambridge University Press.
- Tomaselli, A., Pastore, S., Augliera, P., and Eva, C. : 1992, 'Sismicità dell' Appennino nord-occidentale', *Studi Geologici Camerti 1992/2*, 35–42.
- Tondi, R., de Franco, R., and Barzaghi, R.: 2001, 'Sequential integrated inversion of refraction, wide-angle reflection traveltimes and gravity data for two dimensional velocity structures', *Geophys. J. Int.* **141**, 679–698.
- Zelt, C. A., and Smith, R. B.: 1992, 'Seismic traveltimes inversion for 2-D crustal velocity structure', *Geophys. J. Int.* **108**, 16–34.

PRESENT-DAY SEA LEVEL CHANGE: OBSERVATIONS AND CAUSES

A. CAZENAVE, C. CABANES, K. DOMINH, M.C. GENNERO and C. LE PROVOST
*Laboratoire d'Etudes en Géophysique et Océanographie Spatiales (LEGOS-GRGS)/CNES,
Toulouse, France (anny.cazenave@cnes.fr)*

Received: 26 June 2002; Accepted in final form: 4 November 2002

Abstract. We investigate climate-related processes causing variations of the global mean sea level on interannual to decadal time scale. We focus on thermal expansion of the oceans and continental water mass balance. We show that during the 1990s where global mean sea level change has been measured by Topex/Poseidon satellite altimetry, thermal expansion is the dominant contribution to the observed 2.5 mm/yr sea level rise. For the past decades, exchange of water between continental reservoirs and oceans had a small, but not totally negligible contribution (about 0.2 mm/yr) to sea level rise. For the last four decades, thermal contribution is estimated to about 0.5 mm/yr, with a possible accelerated rate of thermosteric rise during the 1990s. Topex/Poseidon shows an increase in mean sea level of 2.5 mm/yr over the last decade, a value about two times larger than reported by historical tide gauges. This would suggest that there has been significant acceleration of sea level rise in the recent past, possibly related to ocean warming.

Keywords: Sea Level, Thermal Expansion, Water Cycle, Global Warming, Climate

1. Introduction

Determination of present-day sea level change is a subject of high interest in the context of the present debate on global climate change. Several components of the climate system indeed contribute to changing the global mean sea level. Processes involved are: change of the ocean volume, in response to variations of sea water temperature and salinity at all depths (the two parameters having opposite effects on sea level), and change of the ocean mass as a result of water exchange with other surface reservoirs (atmosphere, continental waters, glaciers and ice sheets). Modeling these effects is difficult because complex physical mechanisms, not yet fully understood, are involved. Besides, measuring temporal and spatial characteristics of sea level change, at various time scales, is an important goal since it provides boundary conditions to the models. For the past few decades, long-term sea level change has been estimated from tide gauge measurements. From these observations, it has been deduced that the global mean sea level has been risen by about 2 mm/yr since the beginning of the 20th century (see for example Douglas, 2001, for a review). Since the beginning of the 1990s, Topex/Poseidon altimetry is monitoring sea level variations, with high-precision, high spatio-temporal resolution, global coverage of the oceans, and absolute sea level measurements in a



terrestrial reference frame tied to the Earth's center of mass (see Fu and Cazenave, 2001, for a review).

In this paper we discuss processes that are responsible for global mean sea level change on interannual to decadal time scales. Due to recent availability of long time series of global data sets either from direct observations or model outputs, it becomes possible to estimate two important contributions to sea level change: thermal expansion and water mass exchange between oceans and continental reservoirs. We focus on three different time spans. First, the past decade during which Topex/Poseidon altimetry has observed global mean sea level change. Recent availability of global ocean temperature data makes possible to estimate thermal expansion and compare it with Topex-Poseidon observations. Second, the past 20 years for which new global soil water and snow data allow us to estimate the land water mass contribution and compare it to thermal expansion. Finally, the past 40 years which is the period covered by ocean temperature data. For that period, we estimate the thermal expansion contribution and compare it with sea level observations from tide gauges.

2. Sea Level Change During the 1990s

Analyses of Topex-Poseidon altimetry data indicate that in terms of global mean, sea level has been risen by about 2.5 mm/yr since early 1993 (*e.g.*, Nerem and Mitchum, 2001a, 2001b; Cabanes *et al.*, 2001; see Figure 1). Satellite altimetry also allows mapping the spatial characteristics of the observed rate of change (Figure 2). Both geographical distribution and global averages have indicated that altimetry-derived sea level and sea surface temperature trends are highly correlated, which suggests that, at least part of the observed sea level change has a steric (thermal) origin (Chambers *et al.*, 2000; Cabanes *et al.*, 2001).

In a recent study (Cabanes *et al.*, 2001), we quantitatively estimated the thermal contribution to the Topex/Poseidon-derived sea level change using recently released global ocean temperature data down to 3000 m, for the past 50 years (Levitus *et al.*, 2000). These global ocean temperature data consist of $1^\circ \times 1^\circ$ gridded temperature fields, derived from objective analyses of historical temperature profiles, given as yearly means for the upper 500 m for 1945–1998 and as 5-year means for depths down to 3000 m for 1945–1996. We used the yearly mean temperature data (available for the upper 500 m only) to compute the thermosteric sea level (*i.e.*, due to thermal expansion) for comparison with the Topex/Poseidon-derived sea level. We considered the 6 years of overlap with the Topex/Poseidon data, *i.e.*, 1993–1998. For purpose of comparison, we also computed yearly averages of the Topex/Poseidon-derived sea level and averaged the thermosteric sea level between 60°N and 60°S as for the altimetry data. Over the years 1993–1998, the thermosteric sea level rise for 1993–1998 amounts to 3.1 ± 0.4 mm/yr. This value is in very good agreement with the 3.2 ± 0.2 mm/yr observed (*i.e.*,

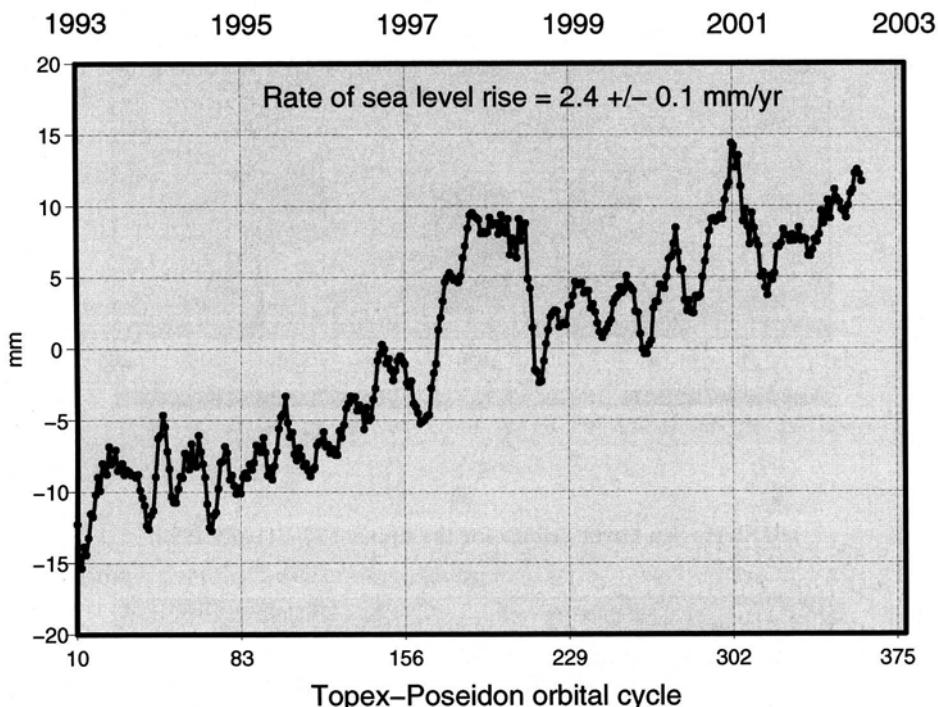


Figure 1. Global mean sea level curve from Topex/Poseidon altimetry for 1993–2001.

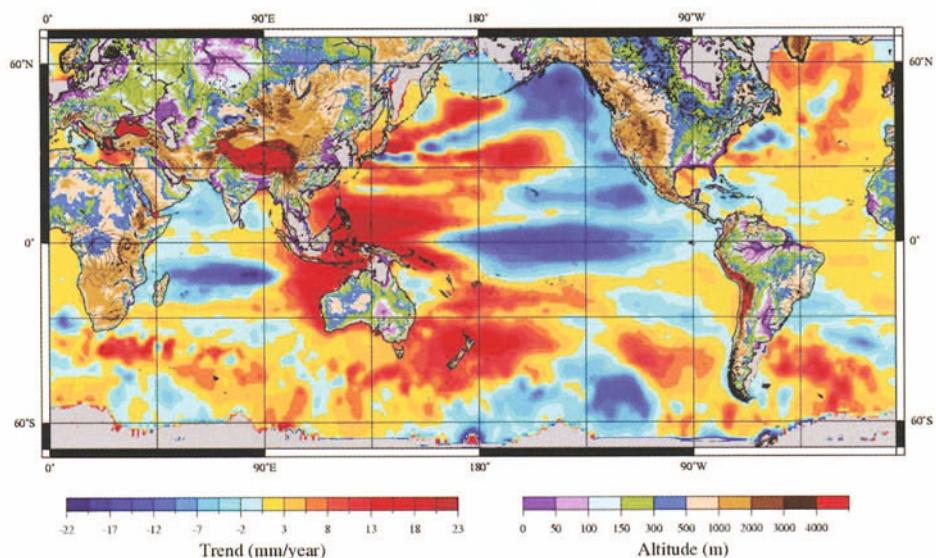


Figure 2. Geographical distribution of sea level trends from Topex/Poseidon for 1993–2001.

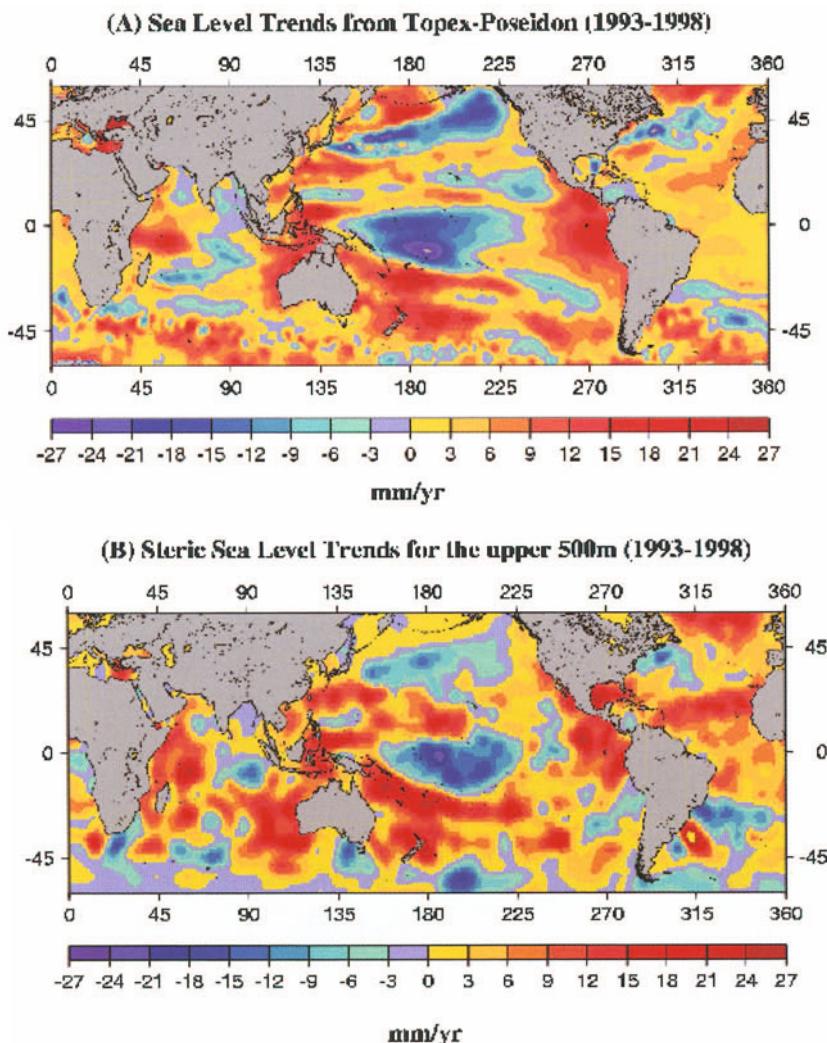


Figure 3. (a) Sea level trends map from Topex/Poseidon for 1993–1998 (b) Steric sea level trends map for 1993–1998, computed from the Levitus *et al.* (2000) temperature data set down to 500 m

Topex/Poseidon-derived) rise over the same time span. The residual sea level (observed minus thermosteric sea level) presents a small, not significant, trend of 0.2 ± 0.2 mm/yr. Figures 3a and 3b present trend maps for 1993–1998 of the Topex/Poseidon-observed and thermosteric sea level, respectively. The agreement between the two maps is striking, in particular in the tropics and Northern Hemisphere. We note in particular that the thermosteric trend map reproduces very well the Eastern Pacific sea level rise associated with the 1997–1998 ENSO event, as well as the Western Indian ocean rise. Similarly sea level patterns in the equatorial and Northern Atlantic are well reproduced both in shape and magnitude by the

thermosteric map. Some discrepancy is observed in the southern oceans where the positive trends observed by Topex/Poseidon are larger than the thermosteric trends, a likely consequence of sparse temperature data coverage in remote southern regions. In addition to uncertainties in observed and computed sea level rise, differences between the two maps may represent three unknown contributions: deep (500–3000 m) thermosteric contribution, water mass addition to the oceans due to exchange with atmosphere and continents (including ice caps), and halosteric (*i.e.*, due to salinity change) contribution. As we will see in section 3, the land water mass contribution is small and of the right order of magnitude to explain the residual trend of 0.2 mm/yr. This suggest that remaining contributions (ice caps melting, salinity and deep ocean effects) are either negligible or cancel each other in terms of global mean sea level. It is worth mentioning that the halosteric contribution, not estimated here, is likely to be small. Antonov *et al.* (2002) have indeed shown that in terms of global average, the halosteric contribution is an order of magnitude less than the thermosteric sea level change. The quantitative comparison presented here shows that, for the past recent years, warming of the upper oceans almost fully accounts for the global mean sea level rise observed by Topex/Poseidon. Thus, the sum of all other climatic contributions (water mass exchange with atmosphere and continental reservoirs, ice caps mass balance, as well as deep ocean below 500 m thermal effects and salinity contribution), is not expected to exceed a few percents of the observed sea level rise over 1993–1998.

3. Land and Water Mass Contribution for 1981–1998

The chapter on sea level change (Church *et al.*, 2001) of the 3rd assessment report of the IPCC (Intergovernmental Panel on Climate change), re-estimates the various factors that have contributed to the 20th century sea level rise. The largest contribution (~ 0.5 mm/yr sea level rise) arises from thermal expansion, *i.e.*, warming of the oceans. This estimate is based on a compilation of results from a variety of coupled atmosphere ocean global circulation models (AOGCMs), as well as from regional and basin-scale *in situ* observations. Melting of alpine glaciers produces 0.25 mm/yr sea level rise. Estimate of the Greenland and Antarctica mass imbalance (accounting for a long-term readjustment since the last glacial maximum plus a present-day climate-related response) contributes in the range –0.2 to 0.4 mm/yr. The most uncertain contribution reported by Church *et al.* (2001) is the change in terrestrial water storage that results from human activities (underground mining, building of dams and artificial reservoirs, diversion of rivers for irrigation, deforestation, urbanization, etc.). The IPCC report proposes a value in the range of –1.1 to + 0.4 mm/yr, an estimate based on updated analyses by Gornitz *et al.* (1997) and Sahagian (2000). According to these authors, a significant contribution comes from extraction of water from underground aquifers that partly reaches the oceans, thus contributes to sea level rise. But impoundment in reservoirs and water removed

from rivers for irrigation which partly infiltrate on soils, may be also important, and of inverse sign.

This estimate does not account for a possible change in the continental water budget due to modifications in the global water cycle. Yet, sea level is expected to change because the amount of water mass exchanged between the oceans and land varies in response to climate change. While feasible at the seasonal time scale (*e.g.*, Chen *et al.*, 1998, Cazenave *et al.*, 2000), on longer time scale (from years to decades), such an estimate has been hindered by the lack of global data of continental water storage. Realistic hydrological models have appeared during the past decades, but it is only recently that some of them have produced information on the interannual variability in land water storage. In a recent study (Milly *et al.*, 2002), we used global land water data from a 'state of the art' global hydrological model (Milly and Shmakin, 2002), over an 18-year time span (1981–1998), to estimate the land water mass contribution to sea level change during the last two decades.

Continental water budget includes water (both liquid and solid) stored in the root-zone (soil moisture) and underground aquifers, in the snow pack and surface water reservoirs (lakes, rivers and wetlands). In terms of water mass change on interannual and decadal time scales, the largest contribution arises from soil moisture, ground waters and snow. The global distribution of these quantities is virtually unknown over most of the world because *in situ* observations do not exist. Remote sensing measurements are presently limited to snow cover extent and its evolution with time (although measurement of snow depth using microwave sensors has been attempted). Thus, global estimates of land water storage essentially rely on hydrological models, either coupled with atmosphere/ocean global circulation models, or forced by observations. The models estimate the variation in land water storage by solving a water budget equation that relates temporal change in land water mass to precipitation, evapotranspiration and runoff. Most recent models are based on sophisticated land surface schemes describing soil/vegetation/atmosphere energy and mass transfers. In a forced mode, these land surface schemes are driven by spatio-temporally varying mass and energy fluxes, the main forcing being precipitation. Model outputs include gridded time series of soil moisture, snow depth, surface runoff, occasionally groundwater.

We used monthly $1^\circ \times 1^\circ$ global grids of soil water, underground water and snow (expressed in unit of water mass per unit horizontal area, *i.e.*, load unit) from the Land Dynamics (LaD) model developed by Milly and Shmakin (2002), and Shmakin *et al.* (2002). For each month of the 18-year time span, we spatially averaged soil water, underground and snow data over the whole land surface, applying a $\cos(\text{latitude})$ weighting. Then we expressed the time series in terms of equivalent sea level by simply multiplying by a factor $-25/0.7$ to account for the relative areas between continents and ocean. The 'minus' sign means that any excess or deficit of land water mass has opposite effect on sea level. The three time series are dominated by an annual signal on which is superimposed interannual variability. Since we are interested here in the latter, we have removed the total seasonal (annual and

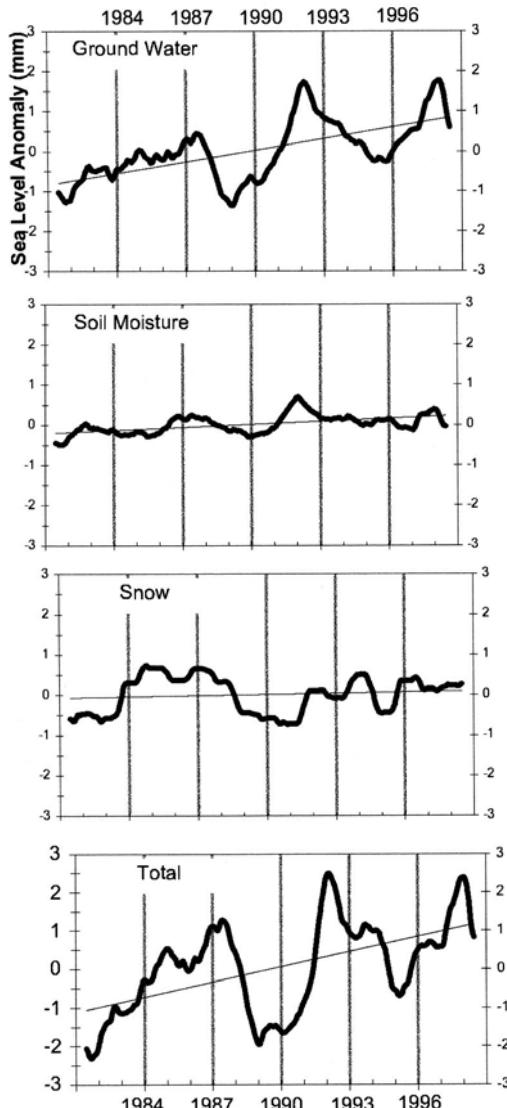


Figure 4. Contribution to the global mean sea level of (a) ground waters, (b) soil moisture, (c) snow and (d) sum of ground waters, soil moisture and snow, for 1981–1998.

semi annual) signal by a least-squares fit and applied a 12-month running mean smoothing to the residuals.

Figure 4 shows the corresponding underground water, soil water, and snow time series, expressed in sea level equivalent. The underground time series shows a clear positive trend on which are superimposed interannual fluctuations. This positive trend amounts to 0.13 ± 0.015 mm/yr mean sea level rise. This corresponds to a linear decrease of underground mass storage during the past two decades. The soil water and snow time series expressed in terms of sea level equivalent

(seasonal term removed, 12-month running mean smoothing) displays a slight positive trend (*i.e.*, sea level rise) but their contributions (0.02 ± 0.005 mm/yr and 0.05 ± 0.005 mm/yr respectively) are very small. We note that during the past two decades, the snow contribution to the global mean sea level was almost negligible. This is unlike the seasonal time scale where snow mass change contributes to about 70% of the observed global mean sea level change, after correction for steric effects (*e.g.*, Cazenave *et al.*, 2000).

As we are interested here on the total land water budget, we added all three contributions (soil moisture, ground water and snow). The positive trend (*i.e.*, sea level rise) of the sum (see Figure 4) amounts to 0.18 ± 0.06 mm/yr. This value is the first direct estimate of land water storage contribution to the sea level rise for the past two decades. In a recent study, Chambers *et al.* (2000) provided an indirect estimate of land water mass contribution to sea level change over a 6-year time span (1993–1998) by subtracting to observed sea level variations from Topex/Poseidon, the steric effects estimated from XBT (expendable bathythermographs) temperature data. These authors interpreted the difference in terms of terrestrial water mass. However their residual curve shows very large interannual variations reaching 20 mm (in terms of sea level) in 2 years. It seems unlikely that such a large change is due to continental waters.

4. Thermal Expansion and Observed Sea Level Rise During the 1955–1995 Time Span

In Cabanes *et al.* (2001), we computed the thermosteric sea level for the period 1955–1996 using the Levitus *et al.*'s 5-year mean temperature data over the depth range 0–3000 m (data over 1945–1955 were not used because original temperature profiles are very scarce). The corresponding thermosteric sea level trend map is shown in Figure 5. The mean thermosteric trend for this 40-year period amounts to 0.50 ± 0.05 mm/yr. Such a value agrees well with the IPCC's estimate based on AOGCMs (see Church *et al.*, 2001). Looking at Figure 5, we note that over the past 40 years, the thermosteric sea level trends display high regional variability, as we previously noticed over the shorter time span.

The most recent estimates of the 20th century sea level rise, based on the longest tide gauge records available (≥ 70 years), report a rate of rise of 1.71 ± 0.55 mm/yr (Douglas, 2001) and 1.84 ± 0.35 mm/yr after correcting for post glacial rebound to account for the visco elastic vertical response of the Earth's mantle to the last deglaciation (Peltier, 2001). These values are more than 2 times larger than the estimate revisited by the IPCC for the total climate contributions (0.7 mm/yr), even if there is small overlap between the respective uncertainties (Church *et al.*, 2001). In order to reconcile the climatic contribution estimates with tide gauge observations, it has been often proposed that thermal expansion based on AOGCMs is underestimated (*e.g.*, Peltier, 2001). However, our calculation based on the global

Steric Sea Level Trends for the upper 3000m (1955-1996)

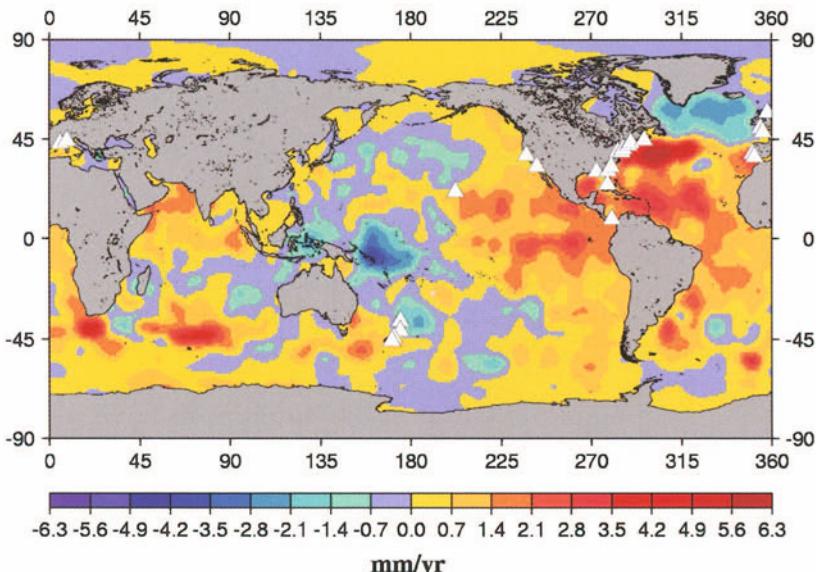


Figure 5. Steric sea level trends map for 1955–1995, computed from the Levitus *et al.* (2000) temperature data set down to 3000 m. The 25 tide gauge sites are indicated by white triangles.

data set of Levitus *et al.* (2000) agrees well with the AOGCMs predictions, an indication that the thermal expansion contribution could hardly account for the difference between the sum of climate-related components and the tide gauge observations. We have seen above that continental water contribution is also much too low to account for the discrepancy, at least for the past two decades. A possible explanation is that the rate of sea level rise of the past century based on tide gauge observations has been overestimated. As is well known, tide gauges suffer from two problems. Their geographical distribution provides very poor sampling of the ocean basins, especially when studying the climatic signal over the past century. Moreover, they measure sea level relative to land, hence record vertical crustal motions that may be of the same order of magnitude as sea level variations. Post glacial rebound is the only vertical motion that is modeled globally and that can be corrected (*e.g.*, Peltier, 2001). However, several tide gauges are located in active tectonic regions. Moreover most islands on which tide gauges are located have a volcanic origin, some of them showing presently volcanic activity. From these two drawbacks, the former is probably the most severe in view of the high regional variability of thermal expansion. Thus any overestimate of the rate of sea level rise would not be surprising. To test this hypothesis, we have computed a pseudo 'global mean' steric sea level by subsampling the global thermosteric sea level grid at locations close to tide gauge sites. For that purpose, we considered 25 out of the 27 stations selected by Douglas (2001) to derive the 20th century sea level

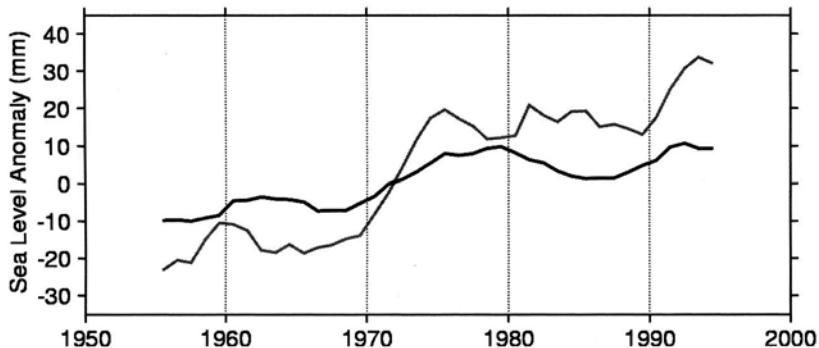


Figure 6. Global mean steric sea level (black curve) for 1955–1995; 'pseudo global' mean steric sea level sub sampled at the 25 tide gauge sites (grey curve) for 1955–1995.

rise. The 25 tide gauges were grouped in 9 regions whose location is displayed in Figure 5, superimposed on the geographical distribution of thermosteric trends for 1955–1996 and 0–3000 m depths (see Cabanes *et al.*, 2001, for details). The pseudo 'global mean' thermosteric sea level time series is superimposed to the 'true' global mean in Figure 6. The linear trend of this pseudo 'global mean' amounts to 1.4 ± 0.10 mm/yr, a value more than 2 times larger than the 'true' global mean trend (of 0.5 ± 0.05 mm/yr). Looking at Figure 5, we note that 23 out of the 25 sites are located in positive trend regions, thus fail to correctly sample the global variation.

We further checked whether the computed thermosteric sea level rise correctly reproduces the tide gauge-derived sea level rise. For that purpose, we considered tide gauges records from the Permanent Service for Mean Sea Level (PSMSL) at the 25 sites. All records were corrected for the inverted barometer response of sea level to atmospheric forcing and post glacial rebound (see Cabanes *et al.*, 2001, for details). The 'average' tide gauge-derived sea level curve over 1955–1996, after performing regional grouping, is shown in Figure 7 superimposed to the 'pseudo global' mean thermosteric sea level. The observed (*i.e.*, tide gauge-derived) mean sea level rise over 1955–1996 amounts to 1.6 ± 0.15 mm/yr, a value that agrees well with the pseudo 'global mean' thermosteric rise (1.4 ± 0.10 mm/yr). The tide gauge-derived sea level curve displays decadal oscillations which origin in non steric and related North Atlantic wind forcing.

In our computation, we did not account for the halosteric component. However, it is unlikely that this neglect will change our conclusion. As mentioned above, in terms of global average, the halosteric component contributes to 0.05 mm/yr sea level rise only (Antonov *et al.*, 2002). Note that this corresponds to a global salinity decrease possibly due to fresh water input to the world oceans. On the other hand, Antonov *et al.* (2002) also showed that the halosteric contribution is quite significant in the subpolar part of the North Atlantic, especially in the Labrador Sea where it nearly counteracts the thermosteric contribution. Thus one may wonder

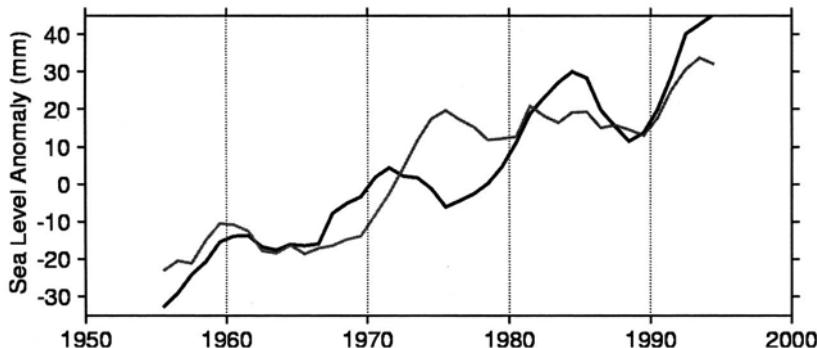


Figure 7. 'Pseudo global' mean steric sea level sub sampled at the 25 tide gauge sites (grey curve) for 1955–1995; Observed mean sea level at the 25 tide gauges sites (black curve) for 1955–1995.

whether neglect of the haline component when computing the steric sea level at the tide gauge sites would change our results. To assess this, we computed, at each of the 25 stations, difference time series of the tide gauge-derived minus thermosteric sea level and fitted a linear trend to these difference time series. The mean value of the difference trends amounts to 0.25 ± 0.14 mm/yr. This value represents an upper bound of the neglected halosteric trend.

The results presented above suggest that the uneven tide gauges distribution and limited sampling available for historical records produce a bias affecting global mean sea level rise estimates. Due to temperature data availability, our analysis is limited to the second half of the 20th century. However, we note that the mean sea level rise computed with this 40-year long tide gauge records agrees well with values based on longer records (Douglas, 2001; Peltier, 2001). We also note that the thermosteric sea level trend of the past 40 years agrees well with AOGCMs computations over the whole 20th century. Thus we believe that our conclusion, *i.e.*, the tide gauge-derived sea level rise for the past few decades is overestimated by a factor of about ~ 2 , holds for the whole 20th century. Such a result would reconcile the estimate of climate-related contributions (on the order of 0.7 mm/yr) as reported by the 3rd IPCC assessment report (Church *et al.*, 2001) with the observed sea level rise.

5. Discussion

On the average, the rate of steric sea level rise for the past 40 years is estimated to 0.5 mm/yr, whereas over the past few years, a rate of 2.5 mm/yr is reported. According to Figure 8, showing the thermosteric sea level from temperature data down to 500 m, the recent rise may just correspond to the rising branch of a decadal oscillation. However the observed sea level from Topex/Poseidon altimetry shows an almost linear rise of 2.5 mm/yr for now 10 years. These observations may suggest that an acceleration took place in the recent past, likely related to warming of the

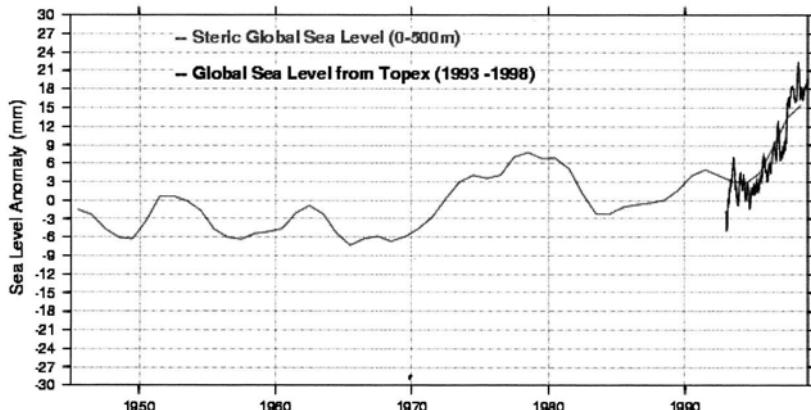


Figure 8. Global mean steric sea level for 1945–1998 computed from the Levitus *et al.* (2000) temperature data set down to 500 m (grey curve); Topex sea level superimposed (black curve).

world ocean. Recent studies based on observations and models indeed indicate that the heat content of the world oceans exhibits a noticeable increase since the early 1970s, possibly related to the increase of anthropogenic gases in the atmosphere (Levitus *et al.*, 2000; Barnett *et al.*, 2001). On going investigations concerning the steric sea level variations over the past 50 years (Cazenave *et al.*, in preparation) indicate that the Atlantic ocean is the main contributor to sea level rise and that the rate of rise increased steeply in the early 1990s. The latter result would suggest that the Topex/Poseidon-derived rate of sea level rise reflects a recent acceleration of thermal origin. However, satellite altimetry and *in situ* temperature data have their own uncertainties and it is still difficult to affirm with certainty that sea level rise is indeed accelerating. According to Nerem and Mitchum (2001a), about 20 years of satellite altimetry data would be necessary to detect, with these data alone, any acceleration in sea level rise.

Future prospects concern both improvements in observing sea level change and climate-related contributions, as well as modeling. On the observational side, important perspective is expected from space data. Firstly, continuous monitoring of global mean sea level variations by altimeter satellites will continue, in particular with the JASON satellite launched in December 2001. The other important perspective is expected from the GRACE mission launched in March 2002. The objective of GRACE is to measure spatio-temporal variations of the Earth gravity field, with an unprecedented resolution ($2^\circ \times 2^\circ$) and accuracy (1 cm on the geoid height) on time scale ranging from 1 month to several years. These observations will give direct insight into the mass redistribution among the surface fluid envelopes. Among the topics studied by GRACE, three concern directly sea level change. Indeed, GRACE will measure continental water mass balance and ice caps mass balance. Over oceanic areas, GRACE will measure ocean mass redistribution, hence after geographical averaging, will provide an independent estimate of the

land water and ice mass contributions to sea level change. Finally, by combining GRACE data with JASON data, it will be possible to separate the thermal from the mass contribution, hence attribute to each process their own effect on the observed sea level rise.

Acknowledgements

This work was supported by CNES and CNRS. We are grateful to R. S. Nerem for having reviewed this manuscript. We would like to also thank him for the extensive talks on this subject we had together during his sabbatical stay at LEGOS.

References

- Antonov, J.I., Levitus, S., and Boyer, T.P.: 'Steric sea level variations during 1957–1994: The importance of salinity', *J. Geophys. Res.*, in press, 2002.
- Barnett, T.P., Pierce D. W., and Schnur, R.: 2001, 'Detection of anthropogenic climate change in the world's oceans', *Science* **292**, 270–274.
- Cabanes, C., Cazenave, A., and Le Provost, C.: 2001, 'Sea level rise during the past 40 years determined from satellite and *in situ* observations', *Science* **294**, 840–842.
- Cazenave, A., Remy, F., Dominh, K., and Douville, H.: 2000, 'Ocean mass variation, continental hydrology and the mass of the Antarctica ice sheet at the seasonal time scale', *Geophys. Res. Lett.* **27**, 3755–3758.
- Chambers, D., Chen, J., Nerem, R. T., and Tapley, B. D.: 2000, 'Interannual mean sea level change and the Earth's water mass budget', *Geophys. Res. Lett.* **27**, 3073–3076.
- Chen, J. L., Wilson C. R., Chambers, D. P., Nerem, R. S., and Tapley, B. D.: 1998, 'Seasonal global water mass budget and mean sea level variations', *Geophys. Res. Lett.* **25**, 3555–3558.
- Church, J. *et al.*: 2001, 'Sea level Change', in J. T. Houghton, Y. Ding, D. J. Griggs, M., Noguer, P. Van der Linden, W. Dai, K. Maskell, and C. I. Johnson (eds.), *Climate Change 2001, The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Program on Climate Change*, Cambridge Univ. Press, Cambridge, pp. 639–694.
- Douglas B. C.: 2001, 'Sea level change in the era of recording tide gauges', in B. C. Douglas, M. S. Kearney, and S. P. Leartherman (eds.), *Sea Level Rise, History and Consequences*, International Geophysics Series, Vol. **75**, Academic Press, London, pp. 37–64.
- Fu, L. L. and Cazenave A.: 2001, *Satellite Altimetry and Earth Sciences; A handbook of Techniques and Applications*, International Geophysics Series, Vol. **69**, Academic Press, San Diego, 463 pp.
- Gornitz, V., Rosenzweig, C., and Hillel, D.: 1997, 'Effects of anthropogenic intervention in the land hydrological cycle', *Global and Planetary Change* **14**, 147–161.
- Intergovernmental Panel on Climate Change (IPCC): 2001, *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Program on Climate Change*, J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. Van der Linden, W. Dai, K. Maskell and C. I. Johnson (eds.), Cambridge Univ. Press, Cambridge.
- Levitus, S., Antonov, J. I., Boyer, T. P., and Stephens, C.: 2000, 'Warming of the world ocean', *Science* **287**, 2225–2229.
- Milly, P. C. D., and Shmakin A. B.: 2002, 'Global modeling of land water and energy balances: 1. The Land Dynamics (LaD) model', *J. Hydrometeorology*, in press.

- Milly, P.C., Cazenave, A., and Gennero, M.C.: 2002, 'Contribution of climate-driven change in continental water storage to recent sea level rise', *Nature*, submitted.
- Nerem, R.S., and Mitchum G.T.: 2001a, 'Sea level change', in L.L. Fu and A. Cazenave (eds.), *Satellite Altimetry and Earth Sciences; A handbook of Techniques and Applications*, International Geophysics Series, Vol. **69**, Academic Press, San Diego, pp. 329–349.
- Nerem, R.S., and Mitchum, G.T.: 2001b, 'Sea level variation from satellite altimetry', in B.C. Douglas, M.S. Kearney, and S.P. Leartherman (eds.), *Sea Level Rise, History and Consequences*, International Geophysics Series, Vol. **75**, Academic Press, London, pp. 121–159.
- Peltier, W.R.: 2001, 'Global glacial isostatic adjustment and modern instrumental records of relative sea level history', in B.C. Douglas, M.S. Kearney, and S.P. Leartherman (eds.), *Sea Level Rise, History and Consequences*, International Geophysics Series, Vol. **75**, Academic Press, London, pp. 65–95.
- Permanent Service for Mean Sea Level: <http://www.pol.ac.uk/pmsl>
- Sahagian, D.: 2000, 'Global physical effects of anthropogenic hydrological alterations: sea level and water distribution', *Global and Planetary Change* **25**, 39–48.
- Shmakin, A.B., Milly P.C.D., and Dunne, K.A.: 2002, 'Global modeling of land water and energy balances: 3. Interannual variability', *J. Hydrometeorology*, in press.

III: OCEAN CIRCULATION

GLOBAL OCEAN DATA ASSIMILATION AND GEOID MEASUREMENTS

C. WUNSCH

*Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, Massachusetts 02139 USA,
(cwunsch@pond.mit.edu)*

D. STAMMER

*Institut für Meereskunde, Universität Hamburg, Remote Sensing, Tropowitzstr.7/III,
22529 Hamburg, Germany, (stammer@ifm.uni-hamburg.de)*

Received: 4 June 2002; Accepted in final form: 18 September 2002

Abstract. Parts of geodesy and physical oceanography are about to mature into a single modeling problem involving the simultaneous estimation of the marine geoid and the general circulation. Both fields will benefit. To this end, we present an ocean state estimation (data assimilation) framework which is designed to obtain a dynamically consistent picture of the changing ocean circulation by combining global ocean data sets of arbitrary type with a general circulation model (GCM). The impact of geoid measurements on such estimates of the ocean circulation are numerous. For the mean circulation, a precise geoid describes the reference frame for dynamical signals in altimetric sea surface height observations. For the time-varying ocean signal, changing geoid information might be a valuable new information about correcting the changing flow field on time scales from a few month to a year, but the quantitative utility of such information has not yet been demonstrated. For a consistent estimate, some knowledge of the prior error covariances of all data fields is required. The final result must be consistent with prior error estimates for the data. State estimation is thus one of the few quantitative consistency checks for new geoid measurements anticipated from forthcoming space missions. Practical quantitative methods will yield a best possible estimate of the dynamical sea surface which, when combined with satellite altimetric surfaces, will produce a best-estimate marine geoid. The anticipated accuracy and precision of such estimates raises some novel modeling error issues which have not conventionally been of concern (the Boussinesq approximation, self-attraction and loading). Model skill at very high frequencies is a major concern because of the need to de-alias the data obtained by the inevitable oceanic temporal undersampling dictated by realistic satellite orbit configurations.

1. Introduction

Physical oceanography and marine geodesy have historically had a long symbiotic history, including periods of discord. The most fundamental relationship has been through the shared problem of defining a marine geoid. For the geodesist, the geoid height is a fundamental quantitative description of the shape of the Earth. To the oceanographer it is a reference surface necessary for computing the oceanic circulation. Other branches of both sciences overlap, including the study of tides,



"mean sealevel", Earth rotation and polar motion, and global and regional sealevel rise and fall.

Many of the issues which still confront us today were apparent in the debate between the late physical oceanographer R. B. Montgomery and the geodetic community about the apparent sealevel slopes along the US (and other regions) coastlines. The seemingly extremely accurate geodetic leveling surveys of that era produced for example, estimates that sealevel sloped upward to the north along the US east coast. Montgomery had pointed out that such a slope would be implausible because it implied that the Gulf Stream had to flow "uphill". However, physical oceanography at that time was widely regarded as an extremely crude subject, not capable of supporting comparison to the extremely accurate leveling surveys. Montgomery's student W. Sturges (Sturges, 1974) revived the debate later on and Montgomery was vindicated (Balazs and Douglas, 1974) when it was finally recognized that systematic errors were producing large-scale discrepancies in the leveling surveys.

The elements of the modern geodetic/oceanographic symbiosis are the same: the seasurface nearly, but not quite, coincides with the geoid; slopes of the seasurface relative to the geoid imply measurable oceanic velocities. Because the seasurface slopes of the seasurface are less than one meter in thousands of kilometers, small errors in estimates of the slopes imply large erroneous oceanic mass and property fluxes. Thus somewhat paradoxically, comparatively crude oceanic circulation estimates can provide relatively accurate estimates of the geoid height slopes. Modern attention to this problem arose with the development of high accuracy satellite altimetry and the various areas of overlap of physical oceanography and geodesy have led to a nearly complete convergence of issues. Here, we will focus primarily on the geoid/circulation problem.

Wunsch and Gaposchkin (1980) described the general problem and laid out the framework for combined estimation of the Earth's geoid and the ocean circulation. More than 20 years later, particularly with the flight of the TOPEX/POSEIDON altimetric spacecraft, enormous progress has been achieved. The problem is worth revisiting as we anticipate the flight of a new generation of spacecraft for determining the gravity field of the Earth by more direct methods (GRACE and GOCE; see Wahr et al., 1998; Drinkwater et al., 2002). In this note we discuss the status of the combined geoid/circulation estimation problem and elaborate on the evolving symbiotic relationship. Ideally, one should use a complete, joint, estimation procedure, but we lack the computational means to carry out the recipe in full. Ocean state estimation has, however, gained a degree of maturity that permits us today to obtain accurate estimates of the ocean circulation. This information can, and should, be used to improve understanding of the geoid.

Before launching into the substance of the problems, it is worth recalling just how far we have come: Consider Fig. 5 in Marsh and Chang (1978). One sees there two of the best marine geoid estimates, from circa 1977, along with the sealevel profile as measured by the GEOS-3 altimetric satellite. Discrepancies of 10 meters exist. Today, with all of the progress in satellite orbit and gravity field

determination, discrepancies on these scales are well below one meter, and are in many places at the centimeter level (see the review by Tapley and Kim, 2001).

Here, we only briefly summarize the basic elements of the geoid/circulation problem. The fundamental relationship derives from the conclusion that the ocean is in near-hydrostatic equilibrium,

$$0 = -\frac{\partial p(\phi, \lambda, t)}{\partial z} - g\rho(\phi, \lambda, t) \quad (1)$$

where g is local gravity, z is the local vertical coordinate, ρ is the oceanic density, ϕ is latitude, λ longitude, and p is the pressure field. Knowledge of the horizontal gradient of the pressure field in the ocean is, in most places and times, sufficient to estimate the flow. At the seafloor, the gradients are easily shown to be

$$\begin{aligned} \frac{\partial p}{a \cos \phi \partial \lambda} &= \frac{g}{a \cos \phi} \frac{\partial [\zeta(\phi, \lambda, t) - N(\phi, \lambda)]}{\partial \lambda} \\ \frac{\partial p}{a \partial \phi} &= \frac{g}{a} \frac{\partial [\zeta(\phi, \lambda, t) - N(\phi, \lambda)]}{\partial \phi} \end{aligned} \quad (2)$$

where ζ is the surface elevation of the ocean relative to the center of the Earth, and N is the geoid elevation, here regarded as time-invariant (its expected time-variability is too small to affect these equations directly).

Although not entirely general, over the great volume of the ocean, the flow field at the seafloor is readily shown to be,

$$\begin{aligned} 2\Omega \sin \phi \rho v(\phi, \lambda, z = 0, t) &= \frac{g}{a \cos \phi} \frac{\partial [\zeta(\phi, \lambda, t) - N(\phi, \lambda)]}{\partial \lambda} \\ &\quad + \varepsilon_\lambda(\phi, \lambda, t) \end{aligned} \quad (3)$$

$$-2\Omega \sin \phi \rho u(\phi, \lambda, z = 0, t) = \frac{g}{a} \frac{\partial [\zeta(\phi, \lambda, t) - N(\phi, \lambda)]}{\partial \phi} + \varepsilon_\phi(\phi, \lambda, t) \quad (4)$$

Ω is the Earth's rotation rate; u, v are the zonal and meridional velocity components. The terms $\varepsilon_{\phi, \lambda}$ are the errors from *both* sides of the equations that appear because the balance of terms is not perfect, and indeed an imbalance is an essential ingredient of the ocean circulation. The imbalances include non-linearity, time-dependence, and stresses. Nonetheless, from an observational point of view, attempts to directly measure the deviation from equality have generally failed to emerge from the noise level. The most obvious deviation from balance arises from the visual conclusion that the circulation evolves rapidly in time (see e.g., the oceanic animation at <http://www.ecco-group.org>). Despite this sometimes violent appearing variability on time scales of days and longer, the so-called geostrophic balance underlying (Eqs. 3, 4) remains extraordinarily accurate (with the notable exception of the ocean within about one degree of the equator).

From altimeters, the time-varying component of ζ is today known with overall accuracies approaching about 2 cm. The range of spatial variation of $\zeta - N$, in a

long-term average, is no more than about 2 m (e.g., Wunsch and Stammer, 1998). If the ocean were at rest, one can make an estimate $\tilde{N}(\phi, \lambda) = \zeta(\phi, \lambda)$. Such a geoid would have errors of 2 m at most, and is more accurate than anything that was available 20 years ago; at high wavenumbers (beyond about spherical harmonic degree 20), it remains the most accurate available marine geoid. Alternatively, if N were known perfectly, then Eqs. (3, 4) would produce $u(\phi, \lambda, z = 0, t)$, $v(\phi, \lambda, z = 0, t)$, with an accuracy determined solely by the altimetric error in ζ , and far exceeding our actual present-day knowledge. Coupled with a knowledge of $\rho(\phi, \lambda, t)$, one would have sufficient knowledge to compute the full three dimensional time-evolving ocean circulation from Eqs. (2). In practice, neither (u, v) , nor N is known perfectly and one seeks to estimate them jointly, both in the time mean and time-varying elements. The problem thus falls under the general subject of state estimation (or, in meteorological terminology, data assimilation).

2. The State Estimation Problem

In the most general terms, ocean state estimation aims to obtain the best possible description of the changing ocean and the external parameters governing its behavior by forcing the numerical model solutions to be consistent with all observations. This model-data combination, if carried out properly, results in a best-estimate ocean circulation – one that is better than can be obtained from either model or data alone. At the same time, the method also identifies model components that need improvement, including surface forcing fields, and produces guidelines to improved oceanic observing systems. (See Wunsch, 1996; Fukumori, 2001; Stammer et al., 2002a).

Because of the fundamental importance of understanding the present and future states of the ocean, the consortium “Estimation of the Circulation and Climate of the Ocean” (ECCO) was funded under the US National Ocean Partnership Program (NOPP) to obtain, through the application of mathematically rigorous assimilation methods, the best possible dynamically consistent estimates of the ocean circulation, which can serve as a basis for studies of elements important to climate (e.g., heat fluxes and variabilities). The ECCO consortium includes efforts at the Massachusetts Institute of Technology (MIT), the Jet Propulsion Laboratory (JPL), and the Scripps Institution of Oceanography (SIO). The resulting model-based syntheses and analyses of the large-scale ocean data set will enable a complete dynamical description of ocean circulation, including aspects that are not directly measured such as insights into the natures of climate-related ocean variability, major ocean transport pathways, heat and freshwater flux divergences (similar for tracer and oxygen, silica, nitrate), location and rate of ventilation, and of the ocean response to atmospheric variability.

Mathematically rigorous data assimilation is most commonly formulated as a least-squares problem in which an objective, or cost, function is minimized subject to, data and model dynamical constraints:

$$J = \sum_t (\mathbf{y}(t) - \mathbf{E}(t) \mathbf{x}(t))^T \mathbf{R}^{-1}(t) (\mathbf{y}(t) - \mathbf{E}(t) \mathbf{x}(t)), \quad (5)$$

where $\mathbf{y}(t)$ are observations distributed in space and time, $\mathbf{x}(t)$ is the model state, $\mathbf{E}(t)$ is an “observation matrix” that computes the model estimate of the observations – here assumed to be a linear combination of state vector elements. $\mathbf{R}(t)$ is the error covariance of the observations. $\mathbf{Q}(t)$ is the model-error covariance. The model, in the form,

$$\mathbf{x}(t+1) = \mathcal{F}[\mathbf{x}(t), \mathbf{q}(t), \mathbf{u}(t), \varepsilon(t), t], \quad (6)$$

is the (discrete-time) temporal evolution equation. Here, \mathbf{q} , \mathbf{u} are the known and unknown, boundary conditions and problem parameters, respectively, and $\varepsilon(t)$ is the model error. It is assumed that $\langle \varepsilon(t) \varepsilon(t)^T \rangle = \mathbf{Q}(t)$, the model-error covariance, is known at least approximately. The model can be imposed upon the objective function J , either by using Lagrange multipliers, or in an “unconstrained optimization” form, using $\mathbf{Q}(t)$ as a weighting matrix in a penalty-function type of formulation in which J would be modified to,

$$J' = \sum_t [(\mathbf{y}(t) - \mathbf{E}(t) \mathbf{x}(t))^T \mathbf{R}^{-1}(t) (\mathbf{y}(t) - \mathbf{E}(t) \mathbf{x}(t)) + \varepsilon(t)^T \mathbf{Q}^{-1} \varepsilon(t)] \quad (7)$$

(This particular J' assumes that it makes sense to minimize the weighted sum of model and observational error; it is not the most general possibility.) The final solution is essentially a weighted least-squares fit of the model to the data with appropriate weights for both. Given the data and a model, the prescription of a priori errors associated with data and model constraints ($\mathbf{R}(t)$ and $\mathbf{Q}(t)$) dictates the quality of the assimilation product. The choice of weight matrices renders the solution, $\mathbf{x}(t)$, if it can be found, to be the maximum likelihood estimate for a linear model. As with all such estimation procedures, this one is reduced to a very large minimization problem.

ECCO state estimate computations are based on the MIT GCM (Marshall, et al., 1997); two parallel optimization efforts, the adjoint method (Lagrange multipliers or constrained optimization method) as described in Marotzke et al. (1999), and a reduced state Kalman filter smoother, e.g., Fukumori et al., (1999) are being used. First results of the global ECCO ocean state estimation based upon the method of Lagrange multipliers are summarized in Stammer et al. (2002a,b,c,d) and preliminary results from the sequential (filter/smooth) results are in Fukumori, et al. (1999). Data employed in ongoing synthesis calculations for the period 1992 through 2001 encompass the full WOCE data set and include absolute and time-varying altimetry, monthly mean sea-surface temperature data, WOCE hydrography, XBT,

equatorial moorings (TAO-array) and profiling float (PALACE) temperature profiles, PALACE salinity profiles, mean surface drifter velocities, time-varying US National Center for Environmental Prediction (NCEP) re-analysis fluxes of momentum, heat, freshwater, and scatterometer wind stress fields. Monthly means of the model state are required to remain within assigned bounds of the monthly mean Levitus et al. (1994) climatology. In addition, and very important, the Lemoine et al. (1996) geoid estimate is used directly as the reference surface for the absolute altimetry, along with a full (non-diagonal) error covariance matrix; thus the best a priori existing geoid height estimate is employed along with a vast array of direct oceanographic observations. To bring the model into consistency with the observations, the initial potential temperature (θ) and salinity (S) fields are modified, internal viscosity and diffusion coefficients are estimated and the surface forcing fields are adjusted. Changes in those variables (often referred to as “control” terms) are determined through a best-fit (in the least-squares sense) of the model state to the noisy observations over the full data period. ECCO results are described, and can also be obtained, at <http://www.ecco-group.org>.

State estimates such as those being carried out in ECCO are computationally very demanding, involving the equivalent of iterative fitting of the model to the data over the entire data time-span. Sequential and Lagrange multiplier methods differ in the details of the computational overheads, but neither is trivial for models being run even at too-coarse resolution. On the other hand, the computations do represent the direction in which these fields must move: they are the only known ways to combine a complete knowledge of the dynamics with all of the data of any kind, and the efforts to render them more efficient and easy to use are going to continue apace.

A few central problems exist; two of them are related to the issue of errors. The first is the representation of the model error, written above as $\varepsilon(t)$. General circulation models are never accompanied by explicit statements of error, and of course, any given model will have different errors for different resolution and different elements, be it the mean temperature and salinity, or the low frequency wave propagation characteristics, or the structure of annual mean sealevel, among an infinity of other possible outputs. The degree of error, generally unknown therefore, controls the extent to which the state estimate fits the model relative to the data, and can make qualitatively important differences to the solution. Model errors are also of many different types, involving internal parameters of the model (mixing coefficients and the like), initial and boundary conditions, lack of resolution vertically and horizontally, in the specification of bottom topography and sidewall conditions, etc. Most existing model-error estimates are little more than guesses, and it is a high priority to learn how to represent model errors quantitatively. Some further discussion of some aspects of this problem is provided by Menemenlis and Chechelnitsky (1998).

Even with a fully specified model error, the state estimation error, i.e., the error of the estimate itself, involves the covariance structure of the full state. If the

state vector contains N elements, in principle, at each time step, the second order statistics of the error alone are a matrix of dimension $N \times N$, and it evolves at each time step. Thus the error covariance structure of the mean seasurface involves propagation of model and data errors through a nonlinear system over many model years. The computational load is forbidding, and is behind much of the effort in the JPL-ECCO program to find useful approximations (see Fukumori, 2001). The absence of full error covariances remains the major limitation on the state estimation approach to geoid determination outlined here.

A successful determination of $\mathbf{x}(t)$ permits us to determine the oceanic flow field that is consistent with observations of all types, including purely geodetic ones and to make a best estimate of N that is also consistent with the observations and known oceanic dynamics. In the following, we will first discuss how improved GRACE and GOCE estimates of the geoid field will advance estimates of the ocean circulation. We will subsequently summarize how improved ocean estimates will feed back into the geoid estimation procedure.

3. Impact of the Geoid on State Estimates

It should be clear that improvements in the observations of *any* field leading to better estimates of $u(z = 0, x, y, t)$, $v(z = 0, x, y, t)$ will improve both the geoid, and the ocean circulation. Such observations include those of the geoid itself. New gravity field observations will improve the state estimation in various ways:

1. Improved geoid height fields reduce the error in Eqs. (2) and thus lead directly to improvements in oceanic circulation estimates. Fig. 1 shows the mean sea surface height field and a near-surface flow field, estimated from TOPEX/POSEIDON data relative to the EGM96 geoid model (Lemoine et al., 1996), and which represents the current best-estimate absolute large-scale oceanic surface flow (see Stammer et al., 2002a).
2. Through the geoid height error covariance specification. All estimation problems have solutions whose quality is directly dependent upon the accuracy of the a priori error covariances.
3. Through new observations of bottom pressure at periods from 2 months to the mission lifetime, and which will provide information about the deep, time-varying ocean circulation that is otherwise generally completely unavailable. The degree to which this information represents qualitatively useful new constraints on the time-dependent circulation will only be known when the data are available.
4. Longer missions will be important for understanding secular trends in the ocean circulation. The ocean exhibits what are interpreted to be real trends, both globally (sealevel, temperature), and of opposing signs over large regions. Models also exhibit such trends, but it is difficult to separate real signals from model drifts owing to numerical approximations, and initial and boundary

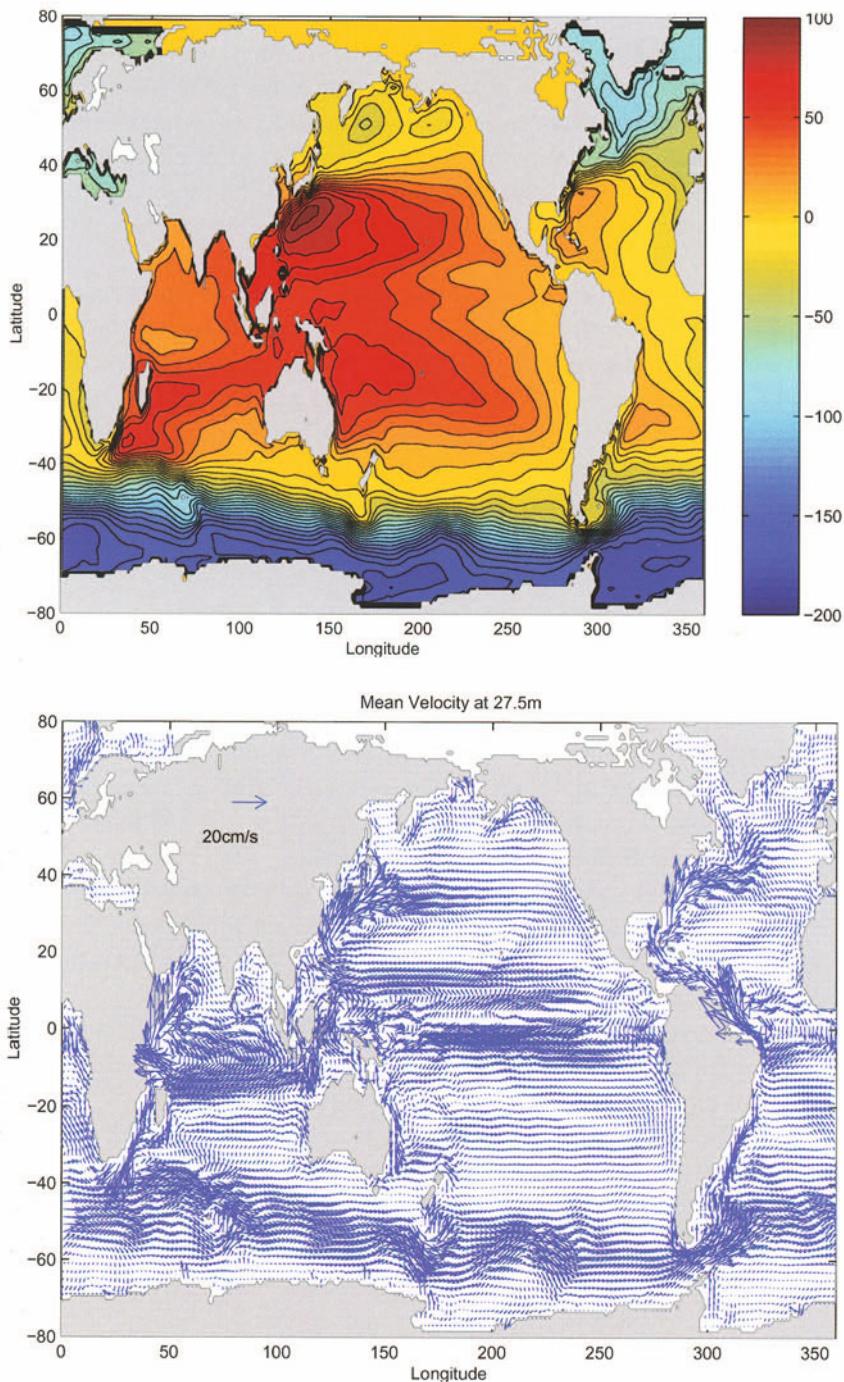


Figure 1. (top) Estimated mean sea surface height field (cm) as it results from the nine-year assimilation period is shown in the upper panel. (bottom) Mean estimated velocity field from 27.5 m (cm/s) from the same period (Stammer et al., 2002a).

condition errors. Gravity measurements should strongly constrain the mass redistribution trends in the ocean permitting separation of real signals from numerical artifacts.

4. Impact of Ocean State Estimates on Geodesy

As already noted, any estimate of oceanic surface flow is equivalent to a knowledge, along with the altimetric measurements, of the geoid height slope. There are several interactions between the state estimates and the geodetic inference problem.

1. Any ocean circulation estimate implies, along with an altimetric surface estimate, a geoid height estimate. Fig. 2 displays the best-estimate geoid from the latest ECCO results (Stammer et al., 2002a). It is visually indistinguishable from other geoids, and so the difference from EGM-96 is shown also. Such ocean-state-estimation geoids will improve as knowledge of the ocean circulation improves by whatever means is available (better theory, better models, better and additional data).
2. Conventional calibration of missions such as GRACE or GOCE is extremely difficult, if not impossible. What is possible are comparisons between GRACE-inferred fields and those independently determined. A major test of the absolute geoid determination from space is provided by the geoid estimate (Fig. 2) from information prior to the mission. A major obstacle here, and one for serious future work, is the great difficulty we have in providing formal uncertainties for the state estimates – the model is non-linear, and of very high dimension; the resulting computational load is currently beyond our capability of handling it.
3. We know (e.g., Stammer et al., 2000), that the gravity missions will alias the surprisingly energetic high frequency barotropic motions in the ocean, because its basic sampling interval is so long (nominally one month). Some of this energy would corrupt the mean state as well as the apparent time-varying geoid height. The best model estimates of that high frequency variability will come from the state estimates and these can be subtracted from the measurements. Model skill will be demonstrated by a measurable reduction in the variance of the resulting corrected fields relative to the uncorrected ones.
4. Oceanic (and atmospheric and core) motions affect the Earth's polar motion and rotation rate (see e.g., Ponte et al., 2001). These motions are of prime concern to geodesists as they affect the reference frames. As ocean models become more skillful, they will permit actual predictions of the polar motion (and as always, measurements of polar motion become useful constraints on the ocean circulation measurements in the general symbiosis).

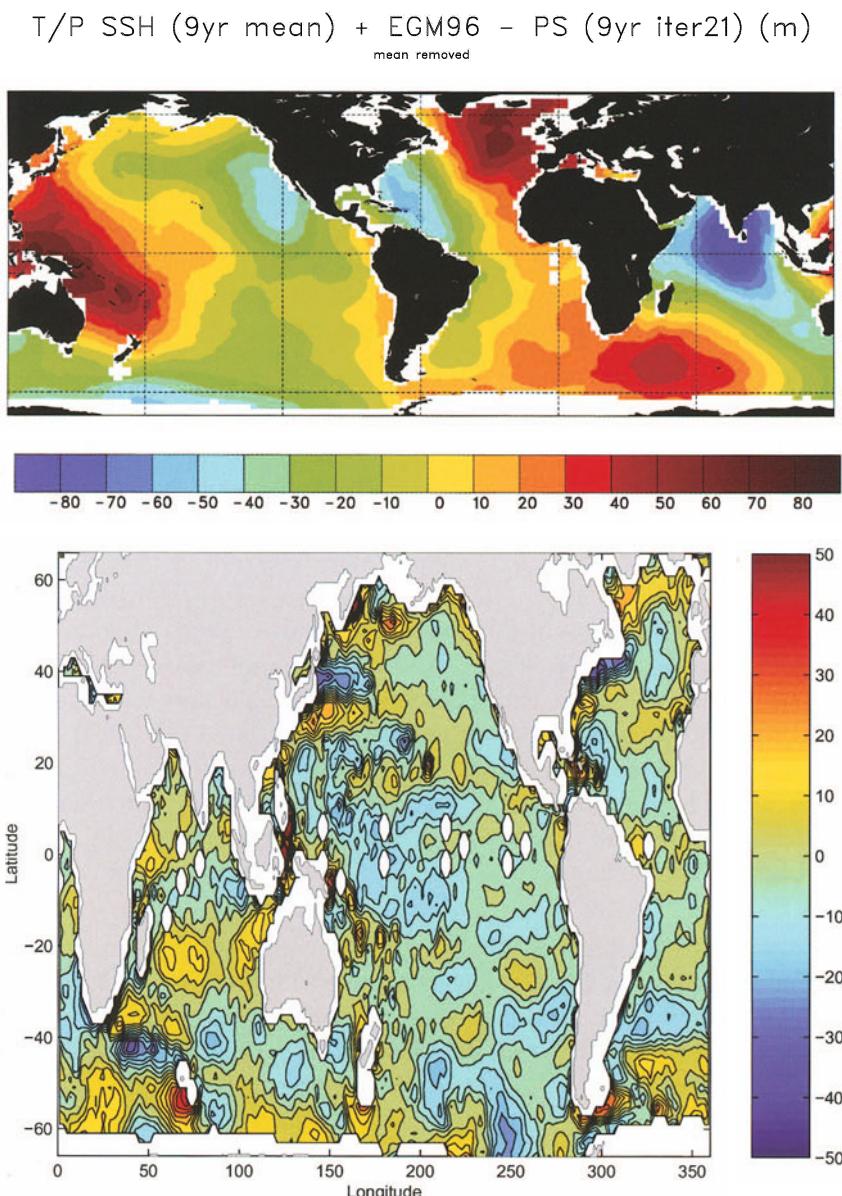


Figure 2. (top) Estimated mean geoid (in meters) as it results from subtracting Fig. 1a from a mean SSH field. (bottom) The estimated mean residual $\bar{\eta}_e - \bar{\eta}_{tp}$ in cm. Note data gaps in the tropical regions due to altimeter track pattern. All data over regions with water depth less than 1000 m were neglected here.

5. Skill

Without formal error estimates, determining the model skill has to be done through a series of comparisons between the model and data, both before and after the data are used as constraints. Note in particular, that the skill in Fig. 2 is tested directly against the in situ oceanic observations by the ability of the model to fit the entire suite of data. Any future geoid, produced by independent means, can be tested against the data by employing it with the model. Geoid height estimates will never be “validated”; rather they will either provide oceanic model estimates consistent with the in situ data, or not so-consistent. (For a useful discussion of the fallacy of “validation”, see Oreskes, et al., 1994.)

As an example of the direct testing of the model against data, we show in Fig. 3 a comparison of two comparatively long records from the Southern Ocean (Spencer et al., 1993) with results at that location from both the unconstrained and constrained models. The unconstrained model clearly tracks the bottom pressure reasonably well. The constrained model does somewhat better, particularly at low frequencies.

A more quantitative comparison of the three estimates for each gauge can be seen in Fig. 4. The spectra of model and data are quite similar at all frequencies. In this preliminary result from the constrained model, most of the increased skill is at low frequencies (the coherence increases in the constrained model there). Note however, that errors of 1 cm of water can create very large oceanic transport errors if they occur over finite distances. This sensitivity is what leads to the expectation that time dependent ocean bottom pressure variations may be powerful constraints on models, and indirectly influence the geoid height estimate.

6. Modeling Issues

Apart from the difficulties mentioned with the absence of model error estimates, the realistic possibility for oceanic bottom pressure measurements from space with precisions approaching or exceeding 1mm of water equivalent (see for example, Wahr et al., 1998; Drinkwater et al., 2002), raises a number of challenging issues for modeling. Hitherto, the need to model, and to use data of this precision, has not been an issue as the data were either non-existent, or much coarser than now anticipated. At the level of precision of GRACE, some conventional approximations used in almost all models begin to fail. These include the so-called Boussinesq approximation that treats the fluid as essentially incompressible (see e.g., McDougall et al., 2002; deSzoeke and Samelson, 2002, and the references there). A perhaps more surprising issue is that discussed by Dewar et al. (1998), who show that the approximation of using depth as a surrogate for pressure in the equation of state of seawater leads to fictitious abyssal pressure gradients of several centimeters of water. Undoubtedly other approximations will have to be dealt with. One example

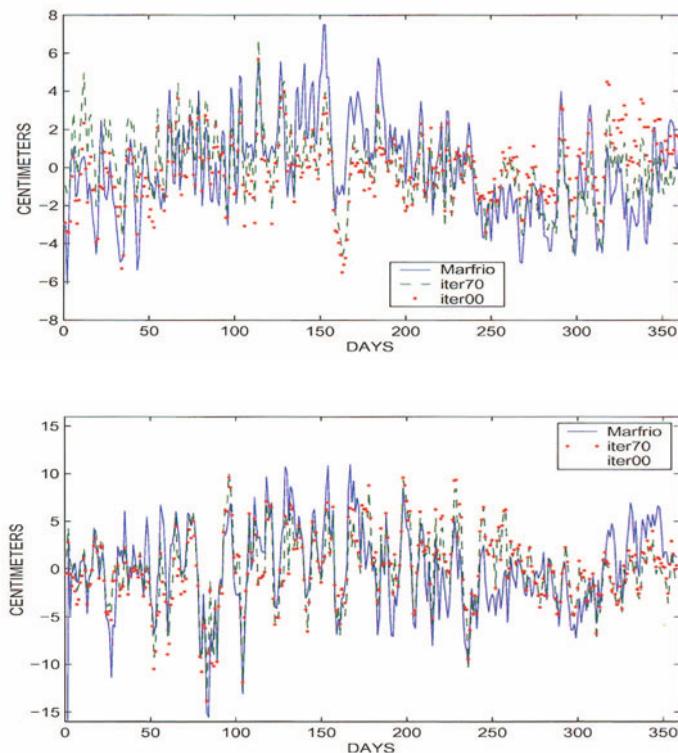


Figure 3. Comparison of ocean bottom pressure variations with an unconstrained model (labelled iter 00), with a constrained one (after 70 iterations). Upper panel is for an instrument at 31°60'S 36°00'W, 2604 m depth, and lower panel is for an instrument at 46°52'S, 52°28'E in 3600 m of water.

is the difficulty in specifying topography in models: the seafloor contains all spatial scales and it is geographically quite inhomogeneous. Typically, topography is averaged over some fixed distance; whether such averages are adequate parameterizations of all of the sub-grid scale topographic effects (scattering) is doubtful. The closure of passages that should be open, by averaging, affects abyssal water mass properties, and hence the gravity field distribution. In many regions, the real topography remains inadequately sampled (further discussion of some elements of the general topographic problem can be found in Losch and Wunsch, 2003).

A somewhat novel issue concerns the extent to which oceanic loads and self-attraction generate measurable effects on the gravity field seen from space. Although these problems are well-known in ocean tidal modelling, they are new in the general circulation context, and are taken up by Condi and Wunsch (2002).

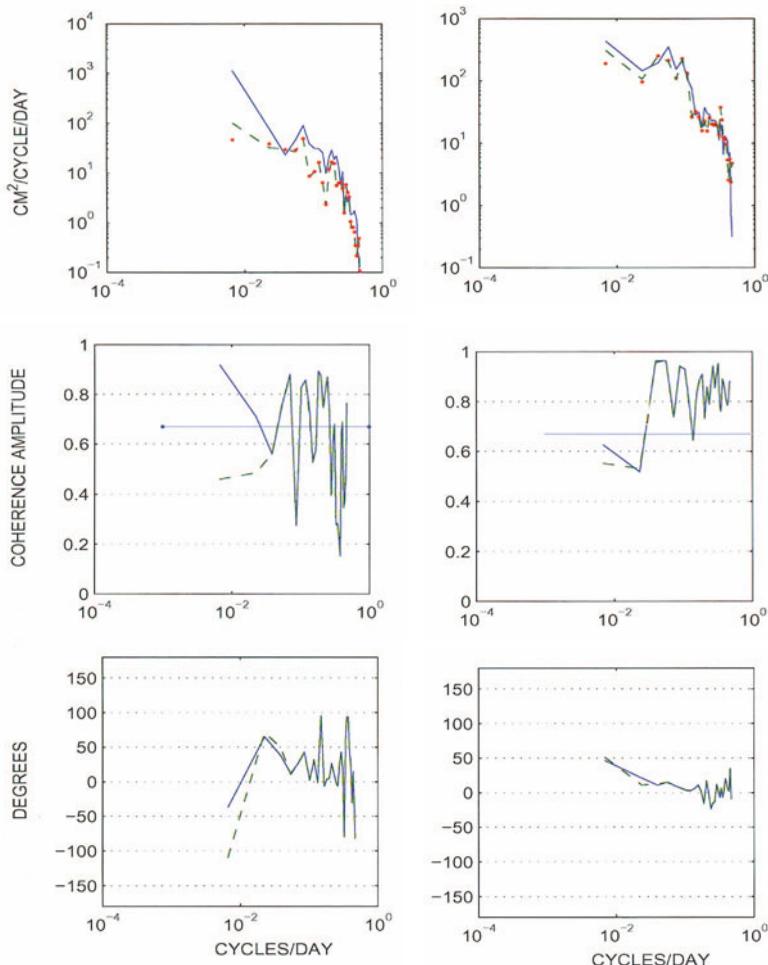


Figure 4. Spectra (upper panels) from the two gauges and the two model estimates shown in 3. Middle panel is the coherence of the records with the two model runs (amplitude) and the lower panels are the coherence phases. An approximate level of no significance for the amplitude is shown.

7. Outlook

The convergence of many aspects of physical oceanography and geodesy can be expected to continue and eventually will mature into a single modeling problem, that will greatly advance both fields. Any information, observational or theoretical, that improves the marine geoid estimates leads to better estimates of the ocean circulation and vice-versa. The most general machinery we have available for using any information of almost any character is that of state estimation in which the data are combined with dynamical and kinematical requirements. Oceanic state estimation has evolved to the stage where both absolute and temporally varying geoid

height information can be combined with oceanic observations, either from space, or in situ, to produce simultaneous best-estimates of both the marine geoid and the ocean circulation. We anticipate continued improvements in the methodologies and as data from new space gravity-measuring missions become available, the community should be able to employ the data nearly routinely. The chief computational problem at the present time is the computational load involved in finding the formal uncertainty estimates for the combined fields; the major conceptual issue is the difficulty in specifying model errors.

A somewhat different role for oceanic state estimates is in the required computation of oceanic time-varying motions at high frequencies so as to reduce the aliasing of missions that necessarily undersample the time-variable ocean. These motions include the tides (which we have not focussed on here), but also the stochastic continuum motions that are most conspicuous at high latitudes. The skill of these models, and their ability to reduce the observed variance, can also be expected to improve as a data stream emerges (Recall Fig. 3.)

Of the two missions specifically dealt with here, GRACE and GOCE, the former provides the novel prospect of observing time-varying ocean bottom pressures, and the latter is expected to provide accurate time average geoids to higher spherical harmonic degree and order than will GRACE. At our present state of experience, it is not clear which, if either, will prove to be the more powerful in setting constraints to improve estimates of either the ocean circulation or geoid height or both.

Acknowledgements

Comments by M. Drinkwater and I. Fukumori were helpful. Computational support from the National Partnership for Computational Infrastructure (NPACI) and the National Center for Atmospheric Research (NCAR) is acknowledged. Supported in part through ONR (NOPP) ECCO grants N00014-99-1-1049 and N00014-99-1-1050, through NASA grant NAG5-7857, through NSF grant OCE 9730071 and through two contracts with the Jet Propulsion Laboratory (958125 and 1205624). This is a contribution of the Consortium for Estimating the Circulation and Climate of the Ocean (ECCO) funded by the National Oceanographic Partnership Program. Also supported in part by the U. of Texas-Austin, Award UTA-98-0222 (GRACE).

References

- Balazs, E. I. and Douglas, B. C.: 1979, 'Geodetic leveling and the sealevel slope along the California coast', *J. Geophys. Res.* **84**, 6195–6206.
- Condi, F. and Wunsch, C.: 2002, 'Measuring gravity field variability, the geoid, ocean bottom pressure fluctuations, and their dynamical implications', *J. Geoph. Res.*, submitted for publication.
- Dewar, W. K., Hsueh, Y., McDougall, T. J., and Yuan, D.: 1998, 'Calculation of pressure in ocean simulations', *J. Phys. Oceanogr.* **28**, 577–588.

- Drinkwater, M. R., Floberghagen, R., Haagmans, R., Muizi, D., and Popescu, A.: 2002, 'GOCE: ESA's First Earth Explorer Core Mission', *Space Sci. Rev.*, this volume.
- Fukumori, I.: 2001, 'Data assimilation by models', in *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave (eds.), Academic, San Diego, 237–266.
- Fukumori, I., Raghunath, R., Fu, L.-L., and Chao, Y.: 1999, 'Assimilation of TOPEX/POSEIDON altimeter data into a global ocean circulation model: How good are the results?', *J. Geophys. Res.* **104**, 25,647–25,665.
- Lemoine, F., et al.: 1997, 'The development of the NASA GSFC and NIMA Joint Geopotential Model', in Segawa et al. (eds.), *Proceedings of the International Symposium on Gravity, Geoid and Marine Geodesy*, International Association of Geodesy Symposia, Vol. 117, 461–469, Springer-Verlag, Berlin.
- Levitus, S., Burgett, R., and Boyer, T.: 1994, *World Ocean Atlas 1994*, vol. 3, *Salinity*, and vol. 4, *Temperature*, NOAA Atlas NESDIS 3 & 4, U.S. Dep. of Comm., Washington, D.C.
- Losch, M. and Wunsch, C.: 2003, 'Bottom topography as a control variable in an ocean model', *J. Atmos. and Ocean Techn.*, in press.
- Marsh, J.G. and Chang, E.S.: 1978, '5' detailed gravimetric geoid in the northwestern Atlantic Ocean', *J. Mar. Geodesy* **1**, 253–261.
- McDougall, T. J., Greatbatch, R. J., and Lu Y.: 2002, 'On conservation equations in oceanography: how accurate are Boussinesq models?', *J. Phys. Oceanogr.* **32**, 1574–1584.
- Marotzke, J., Giering, R., Zhang, Q. K., Stammer, D., Hill, C. N., and Lee, T.: 1999, 'Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity', *J. Geophys. Research* **104**, 29,529–29,548.
- Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C.: 1997, 'A finite-volume, incompressible navier-stokes model for studies of the ocean on parallel computers', *J. Geophys. Res.* **102**, 5753–5766.
- Menemenlis, D. and Chechelnitsky, M.: 1998, 'Error estimates for an ocean general circulation model from altimeter and acoustic tomography data', *Monthly Wea. Rev.* **128**, 763–778.
- Oreskes, N., Shrader-Frechette, K., and Belitz, K.: 1994, 'Verification, validation, and confirmation of numerical models in the earth sciences', *Science* **263**, 641–646.
- Ponte, R. M., Stammer, D., and Wunsch, C.: 2001, 'Improved ocean angular momentum estimates using an ocean model constrained by large-scale data', *Geophys. Res. Letters* **28**, 1775–1778.
- Spencer, R., Foden, P. R., McGarry, C., Harrison, A. J., Vassie, J. M., Baker, T. F., Smithson, M. J., Harangozo, S. A., and Woodworth, P. L.: 1993, 'The ACCLAIM program in the South Atlantic and Southern Oceans', *Intl. Hydrogr. Rev.* **70**, 7–21.
- Stammer, D., Wunsch, C., and Ponte, R.: 2000, 'De-Aliasing of global high frequency barotropic motions in altimeter observations', *Geophysical Res. Letters* **27**, 1175–1178.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C.N., and Marshall, J.: 2002a, 'The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model', *J. Geophys. Res.*, 107(C9), 3118, DOI: 10.1029/2001JC000888.
- Stammer, D., Wunsch, C., Fukumori I., and Marshall, J.: 2002b, 'State estimation in modern oceanographic research', *EOS, Transactions, American Geophysical Union*, 83(27), 289&294–295.
- Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotzke, J., Adcroft, A., Hill, C.N., and Marshall, J.: 2002c, 'Volume, heat and freshwater transports of the global ocean circulation 1992–2000, estimated from a general circulation model constrained by WOCE data', *J. Geophys. Res.*, 108 (C1), 3007, DOI: 10.1029/2001JC001115.
- Stammer, D., Ueyoshi, K., Large, W.B., Josey S., and Wunsch, C.: 2002d, 'Improving air-sea flux estimates through global ocean data assimilation', *ECCO Report Series, Report 13, pp: 31 (see also <http://www.ecco-group.org/reports.html>)*.
- Sturges, W.: 1974, 'Sea level slope along continental boundaries', *J. Geophys. Res.* **79**, 825–830.
- de Szoeke, R. A. and Samelson, R. M.: 2002, 'The duality between the Boussinesq and non-Boussinesq hydrostatic equations of motion', *J. Phys. Oc.* **32**, 2194–2203.

- Tapley, B. D. and Kim, M.-C.: 2001, 'Applications to geodesy', in *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave (eds.), Academic, San Diego, 371–406.
- Wahr, J., Molenaar, M., and Bryan F.: 1998, 'Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE', *J. Geophys. Res.* **103**, 30,205–30,229.
- Wunsch, C.: 1996, *The Ocean Circulation Inverse Problem*, 442 pp., Cambridge Univ. Press, New York.
- Wunsch, C. and Gaposchkin, E. M.: 1980, 'On using satellite altimetry to determine the general circulation of the oceans with application to geoid improvement', *Revs. Geophys. and Space Phys.* **18**, 725–745.
- Wunsch, C. and Stammer, D.: 1998, 'Satellite altimetry, the marine geoid and the oceanic general circulation', *Ann. Revs. Earth Plan. Scis.* **26**, 219–254.

Address for Offprints: C. Wunsch, Massachusetts Institute of Technology, 77 Mass. Ave., Cambridge, Massachusetts 02139 USA, cwunsch@pond.mit.edu

RESOLUTION NEEDED FOR AN ADEQUATE DETERMINATION OF THE MEAN OCEAN CIRCULATION FROM ALTIMETRY AND AN IMPROVED GEOID

CHRISTIAN LE PROVOST (christian.le-provost@cnes.fr)

and MAURICE BREMOND (maurice.bremond@cnes.fr)

Laboratoire d'Etude en Géophysique et Océanographie Spatiale, Toulouse, France

Received: 17 August 2002; Accepted in final form: 4 December 2002

Abstract. The sea surface topography observed by satellite altimetry is a combination of the geoid and of the ocean dynamic topography. Satellite altimetry has thus the potential to supply quasi-global maps of mean sea surface heights from which the mean geostrophic surface ocean currents can be derived, provided that the geoid is known with a sufficient absolute accuracy. At present, however, given the limited accuracy of the best available geoid, altimetric mean sea surface topographies have been derived only up to degree 15 or so, i.e. for wavelengths of approximately 2000 km and larger. CHAMP, GRACE, and the future GOCE missions are dedicated to the improvement of the Earth's gravity field from space. Several studies have recently investigated the impact of these improvements for oceanography, concluding to reductions of uncertainties on the oceanic flux estimates as large as a factor of 2 in the regions of intense and narrow currents. The aim of this paper is to focus on what are the typical horizontal scales of the mean dynamic topography of the ocean, and to compare their characteristics to the error estimates expected from altimetry and these future geoids. It gives also an illustration of the oceanic features that will be resolved by the combination of altimetry and the GRACE and GOCE geoids. It further reassesses the very demanding requirements in term of accuracy and resolution agreed in the design of these new gravity missions for ocean science applications. The present study relies on recent very high-resolution numerical Ocean General Circulation Model simulations.

1. Introduction

Most changes in ocean surface currents, on time scales of a few days or longer, are in geostrophic balance, if we exclude the near surface directly wind driven Ekman flows (see Gill, 1982; Le Grand, 2002; Wunsch and Stammer, 2002). The observation of the gradients of the sea-surface height above the geoid (i.e. the "dynamic ocean topography") thus provides a direct measure of this geostrophic part of ocean surface currents. Hence the concept of measuring the sea surface topography at the global scale from space by radar altimetry (see among other books Stewart, 1985; Rummel and Sanso, 1993). Although it is the most complex ocean observing system ever designed, satellite altimetry has benefited since the first flight of a satellite altimeter in the early 1970's from impressive improvements in measurement accuracy by three orders of magnitude from tens of meters to a few centimetres. Over the last decade, the scientific analysis of the high quality



data supplied by TOPEX/POSEIDON and ERS missions has allowed a tremendous amount of new results in various branches of earth sciences, and in particular in oceanography (see Fu and Cazenave, 2001 for a recent review).

Satellite altimetry has thus the potential to supply a measure of the absolute geostrophic surface ocean currents, but provided that the geoid is known with a sufficient accuracy. The sea surface topography observed by satellite altimetry is indeed the sum of the geoid and of the ocean dynamic topography. While the typical elevations of the mean ocean dynamic topography are of the order of 0.1 to 1 m at spatial scales of 1 degree and below, the precision of present geoid models is of the order of several tens of centimetres at these spatial scales which is of comparative magnitude to topographic expression of many ocean circulation features (Lemoine et al., 1997; Tapley et al., 2001). This is the reason why satellite altimetry has been primarily used to study the temporal variability of the ocean, through the analysis of only the differences in sea surface height measured along precisely repeating ground tracks or at ground track cross overs, this allows measurement of temporal changes while bypassing the uncertainties in the time-invariant geoid. Up to now, because of the limited accuracy of the best available geoids, altimetric absolute dynamic ocean topography has been derived only up to degree 15 or so, i.e. for wavelengths of approximately 2000 km and larger (Woodworth et al., 1998; Fu and Chelton, 2001). And at these scales, as demonstrated by Ganachaud et al., 1997, Le grand and Minster, 1999, and others, the use of satellite altimeter data does not help to improve the general ocean circulation models derived from in situ measurements. To allow improvements, it was concluded that specifically designed gravity missions meeting very demanding requirements were required (Balmino, G., 1998).

Presently there are two active gravity missions dedicated to the improvement of the Earth's gravity field from space: CHAMP, GRACE, and one is planned in the near future: GOCE. CHAMP will improve our knowledge of the geoid but not to the wavelength range useful for oceanic applications (Reigber et al., 2002). GRACE is especially aimed at monitoring the time variations of the gravity field at wavelengths of 500 km and larger (Wahr et al., 1998). GOCE is a high-resolution gravity field mission that will give access to a new range of spatial scales (100–200 km) of major interest for ocean science (Balmino, G., 1998).

Several studies have recently investigated the impact of the improvements of the geoid expected from these missions in the many fields of concern (see Johannessen et al., 2002). Concerning the ocean, improvements expected in water mass and heat flux estimates have been quantified, following the Ganachaud et al., 1997 study, by Le Grand, 2001 and Schröter et al., 2002 with respectively a finite difference inverse and a box inverse model approach using the expected precision of the GRACE and GOCE gravity missions (see Le Grand, 2002). These studies all conclude to significant improvements in terms of reduction of uncertainties on the estimates of oceanic fluxes that will be as large as a factor of 2. Among other results, they point to the fact that the regions where the impact will be the

largest correspond to intense and narrow currents, as expected from the improved resolution of the missions. Hence the interest of a quantitative analysis of the spatial scales in the mean dynamic ocean topography that will be resolved by combining high precision satellite altimetry and these new geoids (Woodworth et al., 1998; Le Provost and Ponchaut, 1999).

Up to now, our knowledge of the mean dynamic topography is limited. The large variability of the ocean circulation in space and time makes direct measurement of the long-term average oceanic general circulation extremely difficult. The way to obtain the best possible estimate of this ocean state is through the optimal combination of in situ observations and numerical model, as described by Wunsch, 1996. Major progress is on going in the development of advanced ocean numerical modelling and data assimilation, including satellite altimetry (see Griffith et al., 2000; Fukumori, 2001; Wunsch and Stammer, 2002). But the recent solutions obtained for example by Fukumori, 1999 or Stammer et al., 2002 are still to 1° resolution, because of the size of the problem to solve. In term of resolution, Ocean General Circulation Model (OGCM) simulations have themselves long suffered from limitations, leading to a systematic weakness in their representation of the smaller scale features (meso-scale turbulence and frontal structures) of the ocean circulation. But thanks to the increasing power of supercomputers, it has been recently possible to improve the resolution of these OGCMs up to a level never used before (Smith et al., 2000; Chassignet and Garraffo, 2001; Siefridt et al., 2002). These simulations capture now the bulk of the energy in the meso-scale. These simulations thus provide pertinent information on the typical horizontal scales of the main ocean currents and frontal features, as well as the intensity of the associated gradients of the mean sea surface dynamic topography.

The aim of this paper is to give a new focus on the typical horizontal scales of the mean dynamic topography of the ocean, to give an illustration of what will be the oceanic features resolved by the combination of altimetry and the geoids expected to be derived from GRACE and GOCE, and to further reassess the very demanding requirements in term of accuracy and resolution agreed in the design of these new gravity missions for ocean science applications. The present study will rely on the exploitation of these very high resolution simulations above cited.

2. Approach

Due to the large variety of processes in the ocean, it is not an easy task to characterise the horizontal scales of the mean dynamic structures of the ocean circulation. We have applied in the following two approaches inspired from Woodworth et al., 1998.

1. The first approach seeks to compare the scales and level of energy of the sea surface gradients (relative to the geoid) associated with the major features of the ocean circulation to the geoid slope accuracy expected from the GRACE

and GOCE space gravity missions. An automatic quantitative method is applied to extract these numbers from the mean sea surface topography (MSST) issued from three of the highest resolution OGCM's experiments available to date.

2. The second approach consists in filtering the MSST issued from these OGCM experiments with different spatial filter cut-off, i.e. (200 km, 500km, 2000 km). 200 km and 500 km are the expected resolution at the centimetre precision of GOCE and GRACE. The 2000 km cut-off is used to illustrate the present state of knowledge.

These approaches have been applied to the North Atlantic Ocean, where the several most advanced high-resolution numerical simulations have been carried out recently.

3. The High Resolution Numerical Simulations

The MSST of three different models have been investigated in parallel in order to insure that the outcome of the study are only weakly dependent upon the model characteristics, as a result of their similarities (primitive equation models, involving the same approximations, similar resolution, realistic forcing fields) rather than their differences (vertical formulation, sub grid scale parameterisation, origin of their forcing fields). The MSST of these simulations were made available by their authors for this study.

3.1. THE MIAMI ISOPYCNIC COORDINATE OCEAN MODEL (MICOM) HIGH-RESOLUTION (1/12°) ATLANTIC SIMULATION

(Paiva et al., 1999; Chassignet and Garraffo, 2001).

The Mercator grid of the model covers the North and Equatorial Atlantic (28°S to 70°N , including the Mediterranean Sea) with a resolution of $\Delta\lambda = 1/12^{\circ}$ and a latitudinal spacing of $0.08^{\circ} * \cos(\varphi)$ where λ , φ are longitude and latitude respectively. This horizontal resolution corresponds to a grid size of 8 km at 10°N , 7.2 km at 36°N and 4.6 km at 60°N . This model is isopycnic with 16 layers in the vertical. We have had access to five years of a simulation (years 9, 10, 11, 13 and 14) forced by monthly climatological COADS-based forcing including fresh-water flux. We have decided to use the year 14 of the simulation, which is the most realistic one, as a result of a better tuning of the horizontal diffusion parameterisation (Chassignet and Garraffo, 2001). We have thus analysed the mean sea surface topography corresponding to this year (MICOM MSST).

3.2. THE MERCATOR PROTOTYPE ATLANTIC AND MEDITERRANEAN SEA (PAM) HIGH-RESOLUTION 1/15° EXPERIMENT

(Siefridt et al., 2002)

The geographic extends of the model is from 9°N to 70°N, including the North Atlantic and the Mediterranean Sea. The horizontal resolution is a complex function of the location, due to a rotated 1/15° Mercator grid with an "equator" line along 60°W, leading to a variable grid ranging between 5 and 7 km, and a local refinement in the Gibraltar Strait. The vertical discretisation is based on 43 levels, varying in thickness between 6 m at the surface and, for the deeper levels, respectively 300 m for the Atlantic and 200 m for the Mediterranean Sea. The mean sea surface here considered is the mean of 3 years (years 6 to 8) after 5 years of spin up forced by the realistic daily wind and heat flux forcing fields of year 1998 from ECMWF (PAM MSST).

3.3. THE LOS ALAMOS PARALLEL OCEAN PROGRAM (POP) HIGH-RESOLUTION (1/10°) ATLANTIC SIMULATION

(Smith et al., 2000)

The model extends from 20°S to 72.6°N, and from 98°W to 17.2°E. The horizontal mesh is a Mercator grid with resolution of $\Delta\lambda=0.1^\circ$ and a latitudinal spacing of $0.1*\cos(\varphi)$ which correspond to a grid size of 10 km at 10°N, 9 km at 36°N and 6 km at 60°N. The model has 40 levels non-uniformly distributed on the vertical, varying in thickness from 10 m at the surface to 250 m at depth. The mean sea surface topography used in this study has been computed from a 3-year period output of the simulation forced with realistic daily wind stresses and seasonal climatological surface heat fluxes (POP MSST)

4. Typical Characteristics of the Major Fronts

4.1. BASIC PARAMETERS CONSIDERED

An automatic algorithm (Bremond et al., 2002) has been developed to detect the presence of topographic fronts over a given MSST, and determine their characteristics in amplitude and width. The algorithm identifies the presence of a front, the direction of maximum slope and, by fitting a cubic polynomial along the frontal profile, determines the typical sinusoid best fitting to this profile, its amplitude A and wavelength λ :

$$\text{MSST}(x) = A \sin\left(\frac{2\pi}{\lambda}x\right) \quad (1)$$

where x is the abscissa taken along this profile (the sinusoid is centred at the inflexion point M of the profile). We can thus deduce the typical amplitude of

the front: $2A$, and its typical slope at the point M of maximum gradient, defined as:

$$\text{grad}(M) = \frac{2A\pi}{\lambda}. \quad (2)$$

And we can choose as typical width of the front the length scale:

$$d = \frac{\lambda}{\pi} \quad (3)$$

(see figure 1).

As an illustration of the products issued from the application of this algorithm, figure 2 shows the location and amplitude $2A$ of the fronts detected in the PAM MSST, of amplitude larger than 10 cm. The characteristic amplitude of the fronts is given through a colour scale displayed along the mean position of the fronts. Similar maps are available for the display of the typical width d of the fronts. We can easily identify in this figure 2 the location of the major currents. The Gulf Stream (GS) typical amplitude is ranging from 50 cm to more than 1 meter, with secondary fronts south of it, associated to the inertial GS recirculation. East of Newfoundland, several fronts materialise the GS extension which extend to the North East, towards the Færöer-Shetland Channel. West of the Mid-Atlantic Ridge, we can follow the mean flows of the Sub Polar Gyre with the pronounced fronts in the Irminger and the Labrador Seas. In the return flow of the Sub Tropical Gyre South of 40°N , the only major fronts are the ones associated with the Azores Current, the Caribbean flow and the Loop Current in the Gulf of Mexico. East-West fronts of smaller amplitude, associated with zonal jets crossing the North Atlantic Basin, are not displayed on this figure.

When applying the algorithm to the MSST issued from the MICOM and POP simulations, similar maps are obtained, with only second order difference that will be shown in the following. These results thus appear to be robust, weakly dependent upon the differences between the three models used. Consequently, we consider that, for the present purpose, they are a good image of the real North Atlantic MSST characteristics.

4.2. PROCESSING OF THE RESULTS

Our aim is to give a synthetic view of the characteristic amplitude and the corresponding spatial scales of the fronts in the MSST issued from the three simulations POP, PAM and MICOM. This is done in the form of a diagram plotting in abscissa the characteristic width d of each front and in ordinate the corresponding value of the sea surface gradient, which relates to the maximum velocity of the flow along the front, through the geostrophic relation. As an illustration, the results are presented in figure 3 for the GS. It can be observed on the left side of figure 3 that the GS mean path is the same for the three simulations before Cape Hatteras, but

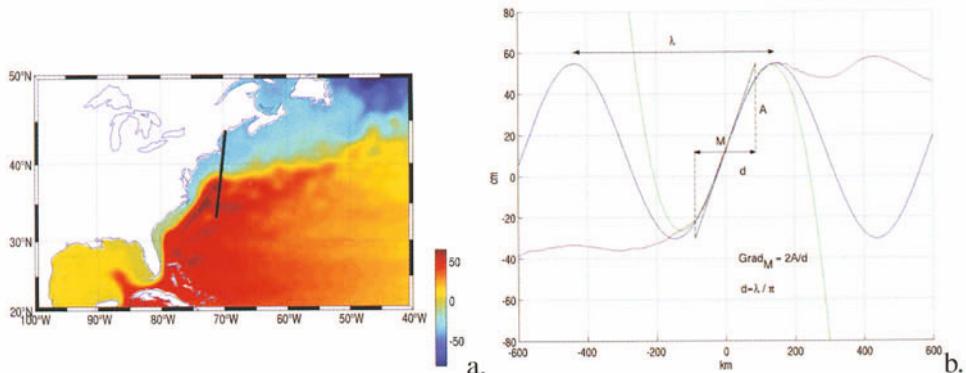


Figure 1. a) Mean Sea Surface Topography from the PAM-05 experiment (scale in cm), with the location of one section scs0 along the maximum slope across the Gulf Stream off Cape Hatteras displayed on figure b). b) Mean Sea Surface Topography along section scs0 (red curve), spline approximation (green curve), sinusoidal approximation (blue curve), and definition of the typical characteristics: grad (MSST) and width d .

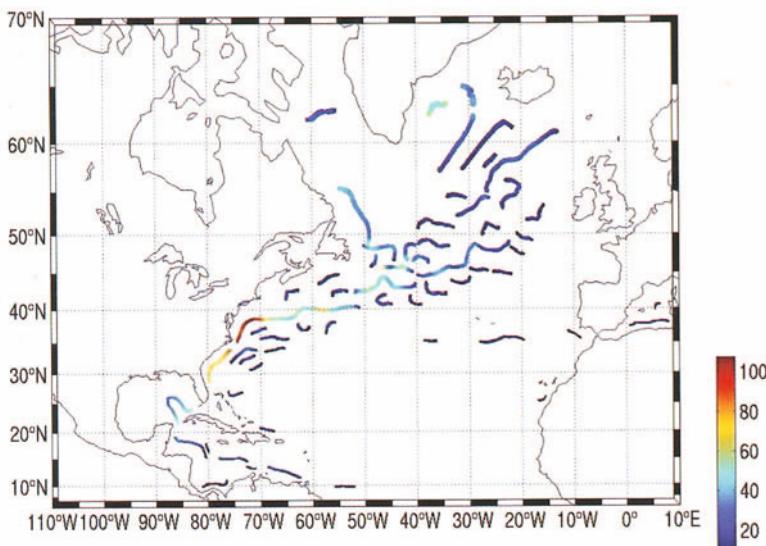


Figure 2. Location and amplitude $2A$ (larger than 10 cm) of the MSST fronts from the PAM-05 experiment. The colour scale is in cm.

is different downstream: this is already known from the authors of these simulations. The behavior of the GS mean trajectory is a very instructive diagnostic of the degree of realism of the simulations. In simulations usually called “eddy 1/4° or lower, the modelled GS had a general tendency to separate far North of Cape Hatteras (see Dengg et al., 1996, for a review). The three simulations considered here have much higher grid resolution. They are now able to better represent this separation, thanks to a better representation of inertial and baroclinic instability

processes (Özgökmen et al., 1997; Smith et al., 2000; Chassignet and Garraffo, 2001). The best fit to in situ observations (Auer, S.J., 1987) is for the MICOM trajectory. The one for PAM slightly overshoots to the North, and the POP GS path is a little too far south. On the right side of figure 3 are plotted the values of the maximum gradient of the MSST along the GS front of the three simulations, as a function of their width d : MICOM in blue, PAM in red, POP in green. The points related to the GS before Cape Hatteras are located in the upper right of the graph. The typical width is 80 to 100 km, in good agreement with what is known from observations. The gradients are ranging from 10^{-5} for MICOM and POP to 6.5×10^{-6} for PAM, which effectively simulate a little weaker GS transport, due to the particular characteristics of the year 98 forcing fields used to run the simulation (Siefridt et al., 2002). For the points representative of the frontal characteristics downstream of Cape Hatteras, d is ranging from 100–150 km to 230 km, with typical gradients decreasing along the GS path from 10^{-5} to 2×10^{-6} , PAM showing again the weaker values.

The same kind of graph has been obtained for the different fronts displayed in figure given as an example for the PAM experiment: the Loop Current in the Gulf of Mexico, the Atlantic Drift west of the GS, the Sub Polar Gyre, the Azores Current, to name only the major currents and associated fronts in the MSST. Figure 4 displays a synthesis of these different plots. It has been obtained by drawing the envelope of the cloud of points related to the same fronts of the 3 MSST. For each current system, the shape of each envelope shows elongated spots that reflect the progressive widening of the flow along its paths downstream, corresponding to smaller gradients/larger front widths. The intense Florida current has a typical width of the order of 100 km and the maximum gradient of the associated fronts is of the order of 10^{-5} . The Loop Current is 100 to 150 km wide, with gradients of 2 to 6×10^{-6} . The GS, as said before, is ranging from 100 km to 230 km, with gradients from 2×10^{-6} to 10^{-5} . The more spread North Atlantic Drift, Sub Polar Gyre current, weaker Azores Currents and the Mid-Atlantic zonal jets, with typical width of 150–250 km, display gradients as weak as 5×10^{-7} .

This diagram is a quantitative confirmation of the highly schematic illustration of the sea surface gradient of the ocean topography given by Woodworth et al., 1998 in their figure 2 for the range of features with characteristics scales smaller than 300 km, i.e. excluding, for the case of the North Atlantic here considered, the larger scale ocean gyres. It is important to note that we are here considering the characteristics of the mean sea surface topography (MSST): it happens that these typical scales coincide with the “meso-scale” which in oceanography designates the oceanic eddy turbulence, hence a possible confusion. The main message of figure 4 (although demonstrated here only through numerical simulation experiments) is that the quasi-stationary “large scale” oceanic circulation contains short scales of the order of 100 km and below. Up to now, these scales are not present in the global inverse solutions of the global mean state ocean circulation, for the many reasons reminded in the introduction. But they need to be in the future, as demon-

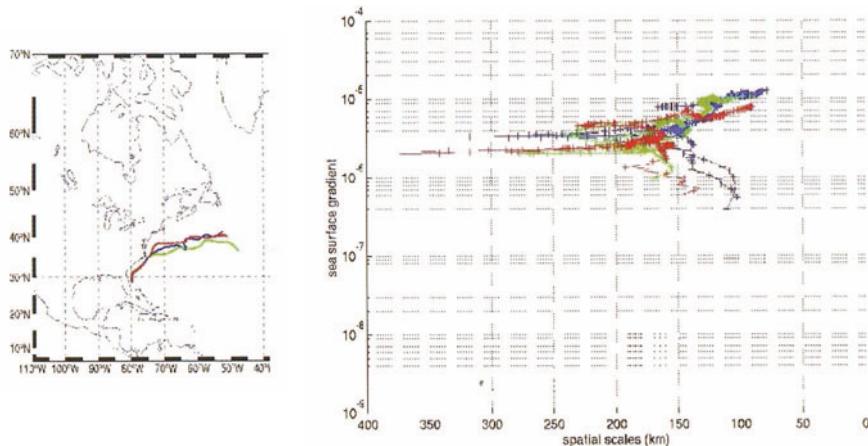


Figure 3. Characteristic MSST gradient and width d along the Gulf Stream, as obtained from the three experiments MICOM (blue), PAM-05 (red) and POP (green).

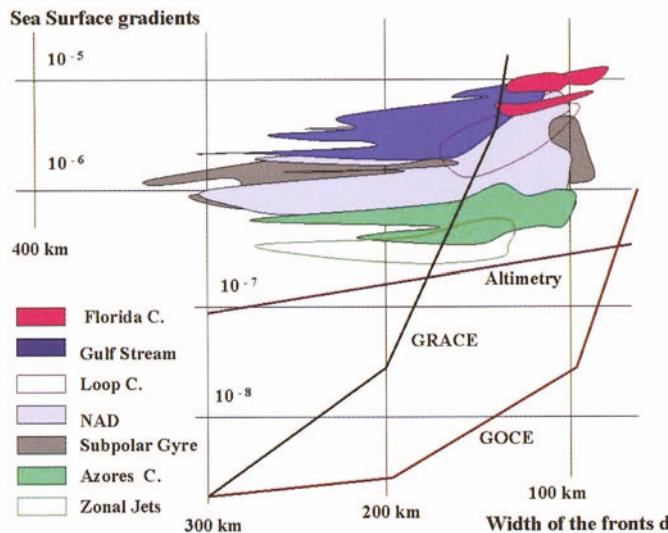


Figure 4. Synthesis of the analysis of the characteristics of the fronts in the MSST simulated by the three experiments MICOM, PAM and POP, along the Loop Current, the Florida Current and the Gulf Stream, the North Atlantic Drift, the Subpolar Gyre, the Azores Current, and typical mid-Atlantic zonal jets. The accuracy estimates for the MSS slope from altimetry, and for the GRACE and GOCE geoids are given for comparison.

strated for example in the present context of geoid improvement and altimetry Le Grand, 2001, Schröter et al., 2002 and others studied the impact of the uncertainty reductions in volume and heat transports which will result from a better geoid. The regions where this impact is the largest correspond to these intense and narrow oceanic currents, with improvements up to a factor of 2.

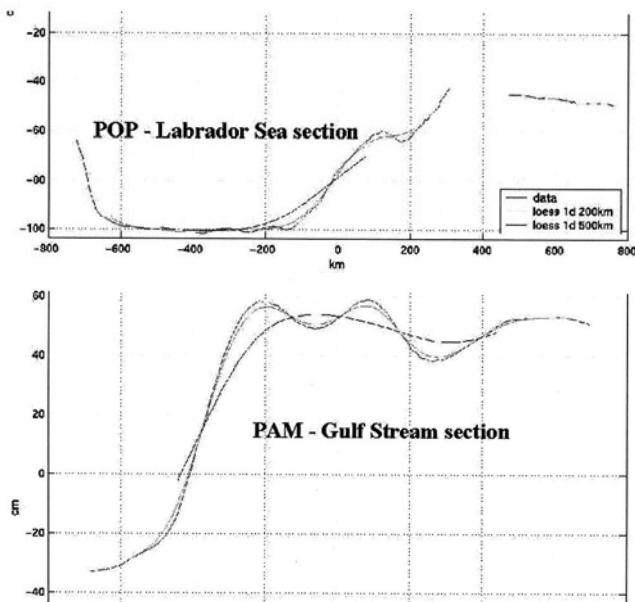


Figure 5. MSST along section sc0 through the Gulf Stream from the PAM experiment (lower panel) and across the Labrador sea from the POP experiment (upper panel) – raw MSST (red), MSST approximation after filtering at 500 km (blue) and 200 km (green) wavelength.

When interpreting these results, we must bear in mind that the concept of “ocean steady state” must be related to a given time scale, because the ocean climate is not steady. The usual time scales considered are generally dependent upon the availability of the data involved in the studies, and the non-resolved time scales appear as noise. For the present study, the MSST are computed from a few years of simulation (one year for MICOM, 3 years for the 2 others). This is somewhat too short, but dictated by the availability of the data, partly related to the very high computer cost of the simulations. One important source of variability is the meso-scale turbulence, and we know from previous studies (see for example Willebrand et al., 2001), that 5 year means are a reasonable requirement for an efficient filtering of this meso-scale turbulence. The dispersion observed in figure 3 between the 3 experiments is certainly partly due to this limitation. But again, for the present purpose, the important result is that the shape of the three “clouds” of points are similar, and of the same order of magnitude.

In figure 4 we have also plotted the typical MSS slope accuracy that we consider to be expected from altimetry in a near future. This estimate could be controversial, and has been addressed in particular in the different impact studies above related (Le Grand, 2001, Schröter et al., 2002). Schrama, 2002 also discusses this question within this volume. Typically for the TOPEX/POSEIDON mission, the single point measurement accuracy of altimeter height is at the 4.0 cm level (Chelton et al, 2001). But we consider here the MSST precision. Taking the T/P altimeter example

with its 10 day cycle, the error along track of the MSST scales down proportional to the square root of the number of cycles involved in the computation of the mean, i.e. 7 mm and 2mm for 1 year and 10 year mean respectively. Besides, as we have pointed out just before, the presence of meso-scale variability, especially in the regions of intense currents, but also elsewhere, could lead to a noise level exceeding the one of the other quantities involved in the computation of the MSST from altimetry. If we adopt a 30cm rms value for the sea level variability in the western boundary currents (Le Traon and Morrow, 2001), then the related errors in the MSST scale as 5 cm and 1.6 cm for respectively 1 and 10 year means. Moreover, we focus here on the maximum slope gradients of the MSSH, which do not coincide with the direction of altimeter satellite ground tracks. We have thus to consider the accuracy of the MSSH objectively mapped at high resolution. This has been addressed among others by Hernandez et al., 2001. Their MSSH is computed on the basis of 7 years of T/P, 5 years of ERS-1/2, 2 years of GEOSAT mean profiles, plus the ERS-1 geodetic data, the mapping (at $1/30^\circ$ resolution) of their error estimate is of the order of 2 to 3 cm, with only a few areas where it goes up to 8 cm. In any case, we consider these values as upper bounds, because it is likely that more advanced data processing and data assimilation schemes will allow reducing them in the near future. The 2 cm uncertainty in the altimetric MSST adopted by Le Grand, 2001 appears thus as a reasonable estimate. The altimetric MSST slope accuracy expresses in figure 4 is then taken as $2\frac{1}{2}\epsilon/d$ with $\epsilon = 2$ cm.

We have also plotted the geoid slope accuracy expected from the GRACE and GOCE space gravity missions on the basis of the predicted cumulative error variances in these geoid (SID, 2000; Visser et al., 2002) and ignoring the geoid slope error covariance. From figure 4, it appears that the signals associated with all the major oceanic features characterised by our analysis are above the estimated noise level of the mean altimetric signal and well above the error level of the future GOCE geoid down to the 80 km half-wavelength resolution. Improvements will also result from the on coming GRACE new geoid; however its noise level will be of the order of (or larger than) the MSST signal for the more energetic part of the main ocean currents. From this graph, it appears that, after both the GRACE and GOCE improved geoid will be available, one limiting factor will be the satellite MSST accuracy, unless the uncertainties on the mean sea surface deduced from satellite altimetry are reduced to the mm level, as mentioned above.

5. Truncated Representation of the Mean Sea Surface Topography

5.1. IMPACT ON TYPICAL SECTIONS

With a geoid accurate at the cm level for wavelength K , it is possible to extract from satellite altimetry the MSST including space scales down to this resolution, provided the uncertainties on the MSST are correctly reduced. The GOCE accuracy

is expected to lead to a geoid at the cm precision at 200 km wavelength resolution. In this section, we will use the MSST supplied by the three above presented numerical simulations MICOM, PAM and POP to illustrate what will be the result of combining altimetry with such an improved geoid. In figure 5, two sections along the maximum slope of the MSST (red curves) are displayed: one normal to the GS front (lower panel) extracted from the PAM-05 simulation, and the other across the Labrador Sea (upper panel) as deduced from the POP experiment.

The GS section is located along the **scs0 section** (see figure 1a) with the US coast on the left of figure 5 (lower panel), and the centre of the Tropical gyre on the right. The steep front associated with the GS is clearly visible. Following the definitions agreed in paragraph 4.1, the typical width of this front is 100 km and its amplitude $2A$ is of the order of 90 cm. South of this front, we observe two oscillations in the MSST which correspond to the westward inertial GS recirculation flow, composed through that scs0 section of two westward veins.

The section in the Labrador Sea starts at Cape Farewell on the Greenland side and ends in the St Laurent Bay. The Labrador Current flows westward at Cape Farewell and southeastward along the coast of Canada. Figure 5 (upper panel) corresponds to a view of the section from the centre of the Labrador Sea, in the Southeast direction. The section is centred on the maximum slope of the current flowing along the Canadian coast. The MSST observed along this section displays effectively a steep slope with amplitude of the order of 30 cm, and typical width of 80 km along the Greenland side and two fronts of respectively 40 cm and 200 km, and 20 cm and 100 km on the Canadian side.

To illustrate what will be the impact of using a GRACE or GOCE type of geoid in extracting a MSST from altimeter measurements, we have filtered these model MSST profiles at 500 km and 200 km cut off scales, which are representative of the expected GRACE and GOCE limit of resolution for a cm level precision. The MSST profiles shown on figure 5 have thus been filtered at these scales by using a Loess filter (Schlax and Chelton, 1992). The result is in blue for 500 km filter, and in green for 200 km. This figure illustrates how well the MSST profiles will be extracted in the GOCE case. For the GRACE case, as expected, this example shows that the shorter wavelength in the MSST associated with the GS recirculation will not be resolved.

5.2. IMPACT AT THE SCALE OF THE NORTH ATLANTIC BASIN ON THE BASIS OF THE PAM MSST

To illustrate what would be the result of combining altimetry with a GOCE type geoid at the scale of the North Atlantic, we have applied a 2D Loess filter with 200 km wavelength cut off to the PAM MSST. We present the result of this filtering in figure 6a. When you compare the filtered topography with the complete one (displayed in figure 1a), we hardly see any difference. To identify these differences, we display in figure 6b the low pass filtered part of the PAM MSST. This illustrates

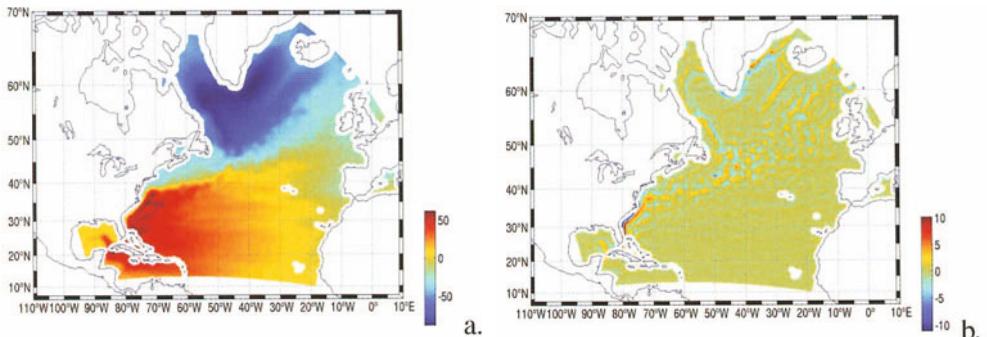


Figure 6. a) MSST filtered at 200 km wavelength from the PAM experiment b) Residual of the MSST filtered at 200 km wavelength for the PAM experiment (difference between the field displayed on figure 1 and the one on figure a).

the part of the MSST that will remain unsolved after to GOCE mission, when simply subtracting the new geoid from the mean sea surface topography built on the basis of altimetry, provided that this altimeter MSST includes all the resolved scales. For the time being this requirement is not met at low and mid latitudes. This topic is addressed in more detail by Chelton, 2001 in the context of high resolution satellite altimetry. As shown also on the two 1D sections of section 5.1, the residual signal is of only a few centimetres, and located mainly along the stronger coastal currents (Florida Strait, GS before Cape Hatteras).

6. Conclusions

We have here assumed that the very high resolution ocean circulation experiments recently run by a few groups of ocean modellers over the North Atlantic Basin give a picture of the dynamic ocean topography which, if not everywhere fully realistic, is certainly very close to reality. To respond to the question “what resolution is needed for an adequate determination of the mean ocean circulation from altimetry and an improved geoid”, we have analysed the characteristics of the MSST issued from very high resolution ocean circulation experiments recently run by a few groups of ocean modellers over the North Atlantic Basin, assuming, in the absence of any measure of this MSST at high resolution, that these simulations give a picture of the dynamic ocean topography which, if not everywhere fully realistic, is certainly close to reality. The analysis has been focused on the topographic fronts associated with the major currents: the GS, the North Atlantic Drift, the Labrador Current, the Azores Current, etc. For each of these features, the typical sea surface gradients and the corresponding spatial scales have been quantified. These gradients are of the order of 10^{-6} to a few times 10^{-7} with typical width d ranging from less than 100 km to a few hundreds km. It has been then possible to relate these characteristics to geoid slope accuracy expected from the GRACE and

GOCE gravity mission and from satellite altimetry. It is here confirmed, through this quantitative approach, that most of ocean current signals in the mean sea surface topography will be correctly resolved by combining satellite altimetry with the future GOCE geoid, provided advanced data processing, the use of data assimilation schemes, and possibly the advent of high resolution satellite altimetry will allow to reduce the uncertainties in the altimetry data. Though these results are limited for the moment to the North Atlantic case, it is reasonable to assume that these conclusions are valid for the global ocean. Unfortunately, no such very high-resolution numerical experiments have been run to the global scale, up to now. Preliminary attempts to run global very high resolution ($1/10^\circ$ and higher) OGCMs are on going in few ocean modelling centres. A quantitative analysis similar to the one presented here will certainly be possible at the global ocean scale when these results will become available.

Acknowledgements

We thank Mark Drinkwater, Jean Michel Lemoine, Pierre Yves Le Traon and the two anonymous reviewers for their help in improving the original manuscript. We also want to thank Eric Chassignet, Rick Smith and the Mercator team for supplying the MSSH of their respective simulations.

References

- Auer, S.J.: 1987, Five-year climatological survey of the Gulf Stream system and its associated rings, *J. Geophys. Res.* **92**, 11,709–11,726.
- Balmino, G., F. Perosanz, R. Rummel, N. Sneeuw, H. Sünkel, and P. Woodworth: 1998, European views on dedicated gravity field missions: GRACE and GOCE *ESD-MAG-REP-CON-001* Earth Sci. Div., Eur. Space Agency, Paris, France.
- Bremond, M. and C. Le Provost: 2002, A front detection method for mean sea surface topography, *LEGOS report*, Toulouse, France.
- Chassignet, E.P., and Z.D. Garraffo: 2001, From Stirring to Mixing in a Stratified Ocean, *Proceedings 'Aha Huliko'a Hawaiian Winter Workshop*, U. of Hawaii, January 15–19, 2001, P. Muller and D. Henderson, Eds., 37–41.
- Chelton, D.B.: 2001, *Report of the High-Resolution Ocean Topography Science Working Group Meeting*, ref 2001-4, Oregon State Univ., Corvallis, 224 pp.
- Chelton, D.B., J.R. Reis, B. J. Haines, L.-L. Fu, and P. S. Callahan: 2001, *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave eds, Academic Press, International Geophysics series, 69, 1–1131.
- Dengg, J., A. Beckmann, and R. Gerdes: 1996, The Gulf Stream separation problem, *The Warmwatersphere of the North Atlantic Ocean*, W. Krauss ed., Borntraeger, Berlin, 253–290.
- Fukumori, I.: 2001, Data assimilation by models, *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave eds, Academic Press, International Geophysics series, 69, 237–266.

- Fukumori, I., R. Raghunath, L.-L. Fu, and Y. Chao: 1999, Assimilation of TOPEX/POSEIDON altimeter data into a global ocean circulation: How good are the results? *J.Geophys. Res.* **104**, 25,647–25,665.
- Fu, L.-L. and A. Cazenave: 2001, *Satellite Altimetry and Earth Sciences*, Academic Press, International Geophysics series, 69, 463 pp.
- Fu, L.-L., and D.B. Chelton: 2001, Large Scale Ocean circulation, *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave eds, Academic Press, International Geophysics series, 69, 133–169.
- Ganachaud, A., C. Wunsch, M.-C. Kim, and B. Tapley: 1997, Combination of TOPEX/POSEIDON data with a hydrographic inversion for the determination of the oceanic general circulation and its relation to geoid accuracy, *Geophys. J. Int.* **128**, 708–722.
- Gill, A.E.: 1982, *Atmosphere-Ocean Dynamics*, Academic Press, New York, 662 pp.
- Griffith, S.M., C. Böning, F.O. Bryan, E.P. Chassignet, R. Gerdes, H. Hasumi, A. Hirst, A.-M. Treguier, and D. Webb: 2000, Developments in ocean climate modelling, *Ocean Modelling* **2**, 123–192.
- F. Hernandez, MH. Calvez, J. Dorandeu, Y. Faugère, F. Mertz, et P. Schaeffer: 2001, Surface Moyenne Oceanique: Support scientifique a la mission altimetrique JASON-1, et a une mission micro-satellite altimetrique, Contrat SSALTO 2945.LOT2-A.1 Rapport Final / Novembre 2001, (CLS/DOS/NT/00.341).
- Johannessen, J.A, G. Balmino, C. Le Provost, R. Rummel, R. Sabadini, H. Sunkel, C.S. Tcherning, P. Visser, P. Woodworth, C. Hughes, P. Le Grand, N. Sneeuw, F. Perosanz, H. Rebhan, M. Drinkwater: 2002, Gravity Field and Steady State Ocean Circulation Mission, Submitted to *Surveys in Geophysics*, July 2002.
- Le Grand, P.: 2002, Impact of geoid improvement on ocean mass and heat transport estimates, this volume.
- Le Grand, P., 2001, Impact of the Gravity Field and Steady State Ocean Circulation Explorer (GOCE) Mission on ocean circulation estimates. Volume fluxes in a climatological inverse model of the Atlantic, *J. Geophys. Res.* **106**, 19,597–19,610.
- Le Grand, P., and J.-F. Minster: 1999, Impact of GOCE gravity mission on ocean circulation estimates, *Geophys. Res. Letters* **26**, 1881–1884.
- Lemoine, F. and 17 others: 1997 The development of the NASA GSFC and NINA Joint Geopotential Model in Proceedings of the International Symposium on Gravity, Geoid and Marine Geodesy, International Association of Geodesy Symposia, Segawa et al., eds., Springer Verlag, Berlin, 117, 461–469.
- Le Provost, C. and F. Ponchaut: 1999 Impact of GOCE mission for ocean circulation study *ESTEC contract 13175/98/NL/GD report*, Eur. Space Res. And Technol. Center, Noordwijk, Netherland.
- Le Traon, P.-Y., and R. Morrow: 2001 Ocean currents and eddies, *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave eds, Academic Press, International Geophysics series, 69, 171–215.
- Özgökmen, T., E.P. Chassignet, and A. Paiva: 1997, Impact of wind forcing, bottom topography and inertia on mid-latitude jet separation in a quasi-geostrophic model *J. Phys. Oceanogr.* **27**, 2460–2476.
- Paiva, A.M., J.T. Hargrove, E.P. Chassignet, and R. Bleck: 1999, Turbulent behaviour of a fine mesh (1/12 degree) numerical simulation of the North Atlantic, *J. Mar. Sys.* **21**, 307–320.
- Reigber, et al.: 2002, A high-Quality Global Gravity Field Model from CHAMP GPS tracking Data and Accelerometry (EIGEN-1S), *Geophys. Res. Letters* **29**.
- Rummel, R., and Sanso F., eds.: 1993, *Satellite Altimetry in Geodesy and Oceanography*, Lect. Notes Earth Sc., 50, Berlin, Springer Verlag, 479 pp.
- Space Research Organization of the Netherlands, Institute for Astronomical and Physical Geodesy, and Delft Institute for Earth-oriented Space Research (SID): 2000, GOCE End to End Performance Analysis, *ESTEC Contract N° 12735/98/NL/GD*, final report.

- Siefriedt, L., Y. Drillet, R. Bourdalle-Badie, K. Beranger, C. Talandier and E. Greiner: 2002, Mise en oeuvre du modèle MERCATOR à haute résolution sur l'Atlantique et la Méditerranée, *MERCATOR Newsletter N°5*.
- Smith, R.D., Maltrud, E.M., Bryan, O. F., Hetcht, M. W.: 2000, Numerical simulation of the North Atlantic Ocean at 1/10°, *J.P.O.* **30**, 1532–1561.
- Schlax, M.G., and D.B. Chelton: 1992, Frequency Domain Diagnostics for Linear Smoothers, *Journal of the American Statistical Association* **87**, N°420.
- Schrama, E.J.O. Error Characteristics estimated from CHAMP, GRACE and GOCE derived GEOIDS and from satellite altimetry derived mean dynamic topography, this volume.
- Schröter, J., M. Losch, and B. Sloyan: 2002, Impact of the Gravity Field and Steady State Ocean Circulation Explorer (GOCE) Mission on ocean circulation estimates 2. Volume transport and heat fluxes across hydrographic sections of unequally spaced stations, *J. Geophys. Res.* **107**, 4-1–4-20.
- Stammer, D., C. Wunsch, R. Giering, C. Eckert, P. Heimbach, J. Marotzke, A. Adcroft, C.N. Hill, and J. Marshall: 2002, The global ocean circulation during 1992–1997, estimated from ocean observations and a general circulation model, *J. Geophys. Res.*, in press.
- Stewart, R.H.: 1985, *Methods of Satellite Oceanography*, Berkeley, Univ. Calif. Press, 360 pp.
- Tapley, B.D. and M.-C. Kim: 2001, Applications to Geodesy, *Satellite Altimetry and Earth Sciences*, L.-L. Fu and A. Cazenave eds, Academic Press, International Geophysics series, 69, 371–406.
- Visser, P.N.A.M., R. Rummel, G. Balmino, H. Sünkel, J. Johannessen, M. Aguire, P.L. Woodworth, C. Le Provost, C.C. Tscherning, and R. Sabadini: 2002, The European Earth Explorer Mission GOCE: impact for the Geosciences *Ice Sheets, Sea Level and the Dynamic Earth Geodynamics Series* **29**, 95–107.
- Wahr, J. M. Molinaar, and F. Bryan: 1998, Time variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.* **103**, 30,205–30,229.
- Willebrand, J., B. Barnier, C. Böning, C. Dieterich, P. Killworth, C. Le Provost, Y. Jia, J.M. Molines, A. New: 2001, Circulation characteristics in three eddy permitting models of the North Atlantic, *Prog. In Oceanog.* **48**, 123–161.
- Woodworth, P., J. Johannessen, P. Le Grand, C. Le Provost, G. Balmino, R. Rummel, R. Sabadini, H. Sunkel, C.C. Tscherning, and P. Visser: 1998, Towards the Definitive Space Gravity Mission, *WOCE Newsletter* **33**, 37–40.
- Wunsch, C. and D. Stammer: 2002, Global ocean data assimilation and geoid measurements, this volume.
- Wunsch, C. and D. Stammer: 1998,, Satellite altimetry, the marine geoid, and the oceanic general circulation *Ann. Rev. Earth. Planet. Sciences*, 26, 219–253.
- Wunsch, C.: 1996, *The Ocean Circulation Inverse Problem*, Cambridge Univ. Press, New York, 442 pp.

ERROR CHARACTERISTICS ESTIMATED FROM CHAMP, GRACE AND GOCE DERIVED GEOIDS AND FROM SATELLITE ALTIMETRY DERIVED MEAN DYNAMIC TOPOGRAPHY

E.J.O. SCHRAMA, (schrama@geo.tudelft.nl)

TU Delft, Department of Geodesy, FMR

Received: 14 May 2002; Accepted in final form: 1 November 2002

Abstract. This paper presents a review of geoid error characteristics of three satellite gravity missions in view of the general problem of separating scientifically interesting signals from various noise sources. The problem is reviewed from the point of view of two proposed applications of gravity missions, one is the observation of the mean oceanic circulation whereby an improved geoid model is used as a reference surface against the long term mean sea level observed by altimetry. In this case we consider the presence of mesoscale variability during assimilation of derived surface currents in inverse models. The other experiment deals with temporal changes in the gravity field observed by GRACE in which case a proposed experiment is to monitor changes in the geoid in order to detect geophysical interesting signals such as variations in the continental hydrology and non-steric ocean processes. For this experiment we will address the problem of geophysical signal contamination and the way it potentially affects monthly geoid solutions of GRACE.

1. Introduction

At the moment there are two active and one planned gravity missions dedicated to the improvement of Earth's gravity field from space. The main goal of the CHAMP mission launched on 15-jul-2000 is to improve the gravity field from GPS tracking data and accelerometer data collected on the spacecraft. The GRACE mission launched on 17-mar-2002 is dedicated to the observation of the Earth's gravity field by measuring the inter-satellite range and velocity between two low-Earth orbiters. For the near future (2006) the European Space Agency (ESA) plans to fly the GOCE gravity gradiometer mission.

The observation of the permanent and/or time variable Earth gravity field is a common goal for all these missions. The static field should be seen as a mean estimate over a defined period. This is nominally 12 months for GOCE and 5 years of GRACE in which case observation of temporal variations in the gravity field is planned on a monthly basis. All mentioned missions thank their existence to a number of applications in Earth sciences, for more details see Dickey (1997) and ESA (1999).

The main motivation for writing this paper is to present an overview on the feasibility of separating the signal of scientific interest from noise sources inherent to all missions. In Sect. 2 we will start with a summary on the expected accuracy



Space Science Reviews **108:** 179–193, 2003.

© 2003 Kluwer Academic Publishers.

level of the gravity field. A part of this discussion is based upon variance propagation techniques whereby it is sufficient to provide a statistical description of the input measurements. Another part is based on results that are currently obtained from the EGM-96 model and more recently derived models. In Sect. 3 we will deal with the problem of estimating the mean oceanic circulation in view of the presence of mesoscale variability, and in Sect. 4 we focus on the separation of scientifically interesting temporal gravity effects in an environment where noise from geophysical correction models is simulated.

2. Accuracy of Gravity Field

2.1. EXISTING GRAVITY SOLUTIONS

Our current knowledge of the Earth's gravity field comes from various information sources that contribute at characteristic wavelengths to a combination solution. To specify the characteristic wavelengths of a particular technique it is helpful to consider the geopotential V whose gradient is the gravitational acceleration. The potential is normally described by a spherical harmonic series expansion:

$$V(r, \theta, \lambda) = \frac{\mu}{a_e} \sum_{lma} \left(\frac{a_e}{r} \right)^{l+1} C_{lma} Y_{lma}(\theta, \lambda) \quad (1)$$

where r, θ, λ represent the radius, latitude and longitude in spherical coordinates, μ and a_e are the gravitational constant and the mean equatorial radius of the Earth. The terms C_{lma} are spherical harmonic coefficients of degree l , order m and parity a while $Y_{lma}(\theta, \lambda)$ are known as spherical harmonics, see also Heiskanen and Moritz (1979). Essentially a gravity model will provide the C_{lma} constants that come together with a set of reference constants like a_e and μ .

Input data for existing gravity solutions comes from two contributing disciplines. One is satellite geodesy where observation of orbit perturbations of one or more satellites are used to model the lower degree and order of the field. Complementary information comes from terrestrial observations by gravimeters, or gravity information inferred from satellite altimetry data or airborne gravity data. Both satellite and terrestrial gravity data form the basis of the EGM-96 model described in Lemoine *et al.* (1998). For EGM-96 several decades of satellite tracking data are used to model the geopotential V . Characteristic of the orbit perturbation analysis is that it loses much of its sensitivity beyond degree and order 70 which corresponds to wavelengths of approximately 600 km on the Earth's surface. The EGM-96 model itself is a least squares inversion procedure that results in the C_{lma} coefficients. Its formal solution is complete to degree and order 360, yet the normal matrix is only full till degree 70 while certain assumptions guarantee an efficient block diagonal structure for higher degrees, see also Lemoine *et al.* (1998).

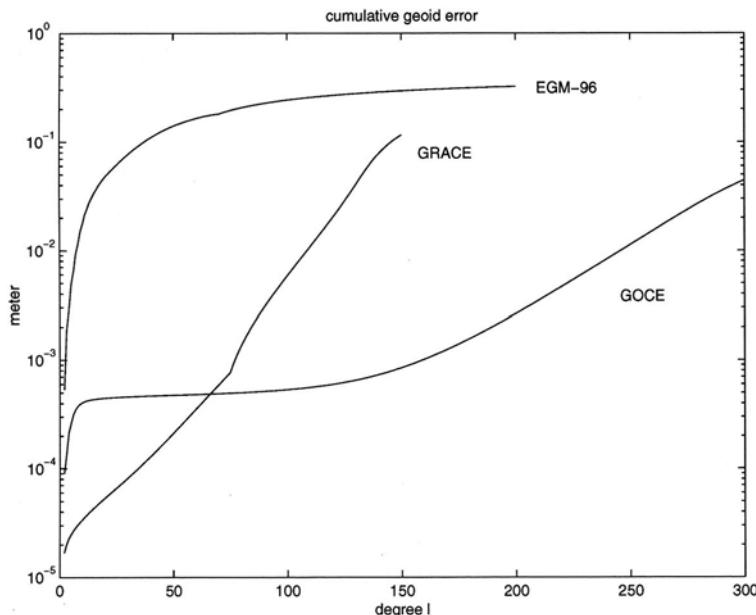


Figure 1. Cumulative geoid errors inferred from the EGM-96 model, worst and best case scenarios for GOCE and GRACE, units: spherical harmonic degree and meters

2.2. NEW GRAVITY MISSIONS

For the three mentioned missions we expect that their observations are added to an existing geoid model. This already happens at the moment of writing where CHAMP data is used in the development of an improved gravity model, cf. Reigber *et al.* (2002). To assess the situation for GOCE we rely on variance propagation such as described in Schrama (1991) for a low earth orbiting gradiometer mission that is tracked by a spaceborne GPS receiver; for GRACE we rely on the technique described in Colombo (1986). The sensitivity of GRACE and GOCE with respect the spherical harmonic coefficients are shown in Fig. 1 and 2. In Fig. 1 we show the cumulative geoid errors of EGM-96, GOCE and GRACE under the assumption that a continuous stream of observation data is collected over a period of 1 year. Specifications with regard to the instrument accuracy and sampling rate etc. can be found in Dickey (1997) and ESA (1999).

From Fig. 1 it can be seen that the EGM-96 geoid error largely exceeds the projected geoid errors that follow from GOCE and GRACE. In Fig. 2 we have introduced worst-best case scenarios for both missions. The worst case scenario for GOCE assumes a gradiometer only solution that is not improved by the GPS tracking data information that is contained in our best case. The worst case scenario of GRACE refers to a 30 day solution, the best case refers to a 5 year solution.

For CHAMP there exists an EIGEN-1S / GRIM5 combination solution that was recently computed by Reigber *et al.* (2002). This is a noticeable improvement

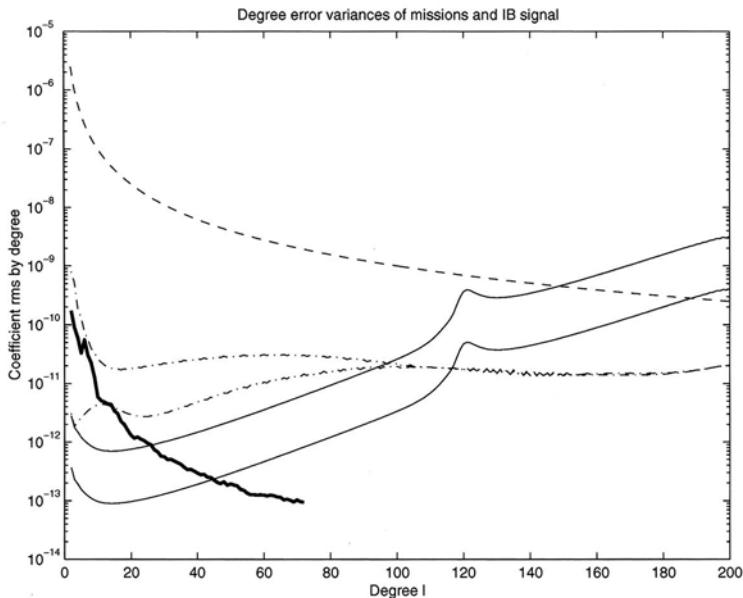


Figure 2. The rms by coefficient by degree for GOCE (dash-dotted line) and GRACE (solid line), mean signal rms according to Kaula's rule of thumb (dashed line) and the air pressure loading signal (thick line). Units: spherical harmonic degree against dimensionless coefficients.

compared to the EGM-96 solution. The error degree variances (not shown in Figs. 1 and 2) for this gravity solution suggest a cumulative geoid error of 1,5 cm at $l = 20$ which is more than factor 2 better.

A caveat emptor on the projections of the GRACE and GOCE geoid errors and degree rms values shown in Figs. 1 and 2 is that we have never inverted a system of normal equations built from real observation data. Instead the data are assumed on a specified nominal orbit and sampled at a regular interval whereby the observation noise power density spectrum is taken according to the instrument specification. In the real world one will face the problem of observation outages (not polar gaps due to sun synchronous orbit etc) or datasets that are shorter than planned etc.

2.3. ERROR CHARACTERISTICS

It is well known that a representation of errors as degree variances corresponds to the assumption of isotropy and homogeneity on the sphere, cf. Moritz (1980). In this case the covariance function for geoid errors that follows from degree variances only depends on the spherical distance and not on the geographical location or azimuth. There exist representations of geoid errors projected from the 70×70 part of the EGM-96 covariance matrix that are not assuming homogeneity or isotropy. In this case the variance at each node in the geoid error map is computed as $\vec{v}^T C_{xx} \vec{v}$ where C_{xx} is the EGM-96 covariance matrix and \vec{v} the corresponding vector of

partial derivatives from the geoid to potential coefficients. The resulting errors maps of the EGM-96 model can be found in Lemoine *et al.* (1998) from which it can be seen that the projected errors represent heterogeneous data quality and coverage as it was used in the model.

Another consideration with regard to error characteristics of gravity models in satellite geodesy concerns a phenomenon known as tailoring. This happens when solutions with or without certain data sub-sets are compared to one another. A good example is the difference between the EIGEN-1S satellite only solution which is exclusively computed from CHAMP satellite tracking data in comparison to the combination solution with the GRIM-5 normal equations, cf. Reigber *et al.* (2002). It turns out that the EIGEN-1S solution adequately models the CHAMP tracking data, yet it does not necessarily result in an accurate description of satellite data not used in the CHAMP solution. Also this is an example where a sub-set solution adequately describes the data that was used rather than to accurately describe the total amount of available data. (in other words, it is a tailored solution).

The same problem could occur if newer GRACE and GOCE data is added to combination solutions. In any case one should avoid the situation that this new gravity model results in less accurate descriptions of existing satellite data and an overly optimistic description of the new input data. The more general discussion is to optimally combine all observational data in one least squares solution and to calibrate this solution. An efficient calibration strategy is described in Lerch (1991), more formal approaches to the same problem can be found in Koch (1986) and Koch (1990).

3. Mean Ocean Circulation Experiment

One of the important objectives of all new gravity missions is to be able to use a very precise geoid surface as a reference for estimating the mean dynamic topography. In this process the mean ocean surface is observed by satellite altimeter systems; it is the difference between the mean ocean surface and the geoid. A relevant question is now whether this procedure will add knowledge to existing hydrographic observations from which the mean ocean circulation field is derived. Moreover we can ask ourselves whether the observation of the mean dynamic topography from satellite altimetry is limited by the existence of meso-scale variability and known oceanographic signals.

Computational schemes that use hydrographic observations to assess the quasi stationary large scale oceanic circulation are described in Wunsch (1996). The observations of horizontal sea water density gradients used in the thermal wind equations provide relations to calculate geostrophic velocities along the vertical relative to a deep reference level where the ocean is assumed in a motionless state. At the ocean surface momentum equations provide an approximation to relate the obtained geostrophic velocities to slopes of the mean dynamic topography.

If satellite altimetry is used then at best a new constraint on the observation of geostrophic velocity can be provided at the ocean surface. In principle these altimetric determined surface velocities can provide extra information on the velocities in the deep ocean or on an improvement on the observed density field. However it is unlikely that detailed variations along the earlier mentioned vertical column can be inferred from this new information. To assess the information content of altimeter data it is desirable to consider a computational scheme whereby the observed density gradients or velocity constraints at the level of no motion and new constraints on the surface from remote sensing technology are assimilated by minimizing a cost function. An example of a recent dynamic topography solution that is based on assimilation of altimetric information with a modification to the cost function concerning the EGM-96 geoid model error can be found in LeGrand *et al.* (2002).

Assessments with respect to the gain in knowledge on the mean dynamic topography and, more importantly, transports associated with this topography utilizing the proposed gravity mission data characteristics are reported in Ganachaud *et al.* (1997), LeGrand (2001) and Schröter *et al.* (2002). Although all papers report differences in the effect of GOCE data on the qualitative improvement of transports in the oceans some results suggest that 30% to 60% seems feasible.

So far all three publications have assumed that there exists a mean dynamic topography observed by altimetry and that it can be compared to an enhanced geoid. Yet the quality label of the mean dynamic topography estimated in this way is certainly not homogeneous due to the presence of mesoscale variability. In this context mesoscale variability is a noise source since it affects the estimation of a mean dynamic topography field from satellite altimeter data. The presence of this noise motivates a study on the sampling problem that arises when we convert the actual sea level into the mean dynamic topography by straightforward averaging.

From Fu *et al.* (1994) it is known that the single point measurement accuracy of altimeter heights over the actual sea level is approximately 5.0 cm. Since that time this figure has somewhat improved due to better orbits, tide models and other instrumental and environmental corrections, yet this doesn't reduce the single point measurement accuracy below the 4.0 cm level. Most of the variance that remains is related to mesoscale variability, ENSO effects and other long periodic signals in the sea level. Noteworthy is extreme mesoscale variability in the western boundary current regions that reaches 30 cm rms, as described in Fu and Cazenave (2001). These high variability regions are interestingly enough directly situated near the steepest gradients in the mean dynamic topography in the western boundary current regions.

It is obvious that the presence of mesoscale variability easily exceeds the noise level of other quantities involved in the computation of a mean sea level. Mesoscale variability is essentially a random process whereby eddies are generated almost at random in western boundary currents. Eddy climatology was studied by many authors for various regions; eddies separate and recycle in western boundary currents, or eddies follow an autonomous path away from the shedding area. Interactions

have been studied with bathymetry, the link to ENSO intensity has been studied, etc. To some extent it is possible to predict the state of the mesoscale eddy field about a month ahead in time as is shown on NRL (2002) where real time altimeter data and thermal image data are assimilated in the Navy Layered Ocean model (NLOM) originally described in Hurlburt and Thompson (1980).

For the moment we assume that the computation of the mean sea level in the western boundary currents will be based on averaging procedures in order to scale down the background noise in altimeter data. The question is now, is this somewhat heuristic approach sufficient to provide the input data for the inversion procedures described in Ganachaud *et al.* (1997), LeGrand (2001) and Schröter *et al.* (2002). In other words, what accuracy level should we assign to mean dynamic topography data computed by straightforward temporal averaging of satellite altimetry data that is not freed-up by subtracting the state of the mesoscale eddy field?

Our heuristic concept is worked out in the following where we consider that the sea level is mapped in a 10 day cycle by the T/P altimeter. If the sea level measurements are averaged over a period of about a year (equivalent to 36 cycles) then the error along a ground track scales down proportional to the square root of the number of cycles involved resulting in mean dynamical topography error of 8 mm, if 10 years of T/P data is used the error will come down to 2.6 mm. Our simplified model shows that the situation will only deteriorate in high energetic mesoscale variability regions since straightforward pointwise averaging will result in 5 cm and 1.6 cm noise when applied over a period of 1 or 10 years respectively. These values are still a factor 10 or more higher than the projected 2 mm geoid error claimed for instance for GOCE at $l = 180$, see also Fig. 1.

Therefore it is likely that more advanced data processing techniques should be used to estimate a more realistic mean climatological current path in western boundary regions from remote sensing data. A discussion on this point is beyond the scope of this paper, we only want to mention the existence of more advanced data assimilation schemes that rely on the combined input of multiple altimeter systems and infrared imaging systems as described in Lillibridge *et al.* (2000) or shown on NRL (2002). Other options suggested by LeGrand (2001) are to scale down the contribution of altimeter information in the definition of the cost function to properly balance the information contained in remote sensing information to hydrographic observations especially in Western boundary regions.

4. Observation of Temporal Gravity

The observation of temporal changes in the gravity field is one of the exiting new ideas that can hopefully be realized by the GRACE mission. This mission should be able to perform this task thanks to its extreme sensitivity below degree and order 50, see Fig. 2 and Wahr *et al.* (1998). The term “temporal gravity” should be interpreted as the gravitational effect of mass changes due to geophysical pro-

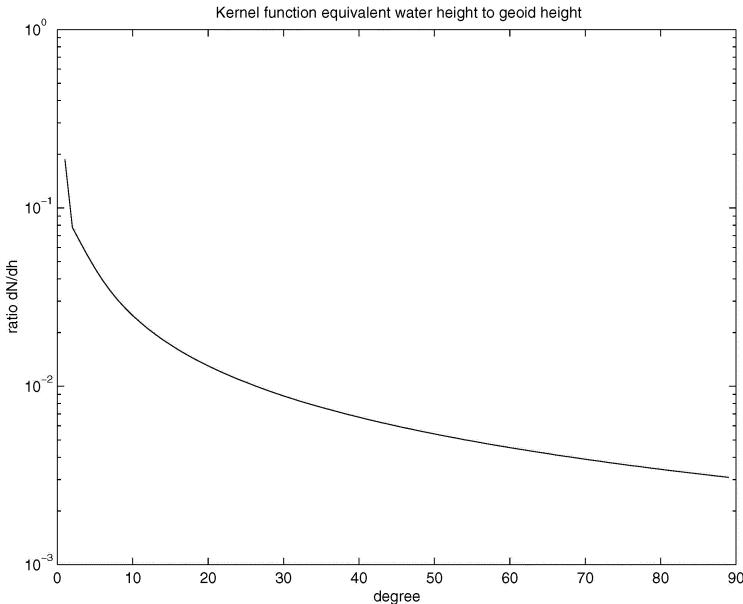


Figure 3. Spectral properties of the convolution operator that converts equivalent water height on the sphere to a self attraction geoid effect.

cesses. Several candidates are mentioned in Dickey (1997), the list includes mass changes as a result of ocean tides, atmospheric effects, continental hydrology, ice volume changes, sea level changes unrelated to temperature, post glacial rebound, earthquakes, mantle convection, tectonic processes, and processes in the Earth's core and mantle.

In this section we will start with the convolution operator that provides the relationship between the self gravitation geoid effect ΔN as a result of the presence of a mass layer with density ρ_w condensed at the Earth's surface modelled as a sphere with radius r_e . The mass layer is expressed in spherical harmonics:

$$\Delta h(\theta, \lambda) = \sum_{lma} \Delta H_{lma} Y_{lma}(\theta, \lambda) \quad (2)$$

and the relation between ΔH and ΔN is:

$$\Delta N_{lma} = \frac{3(1 + k'_l)(\rho_w / \rho_e)}{(2l + 1)} \left(\frac{r_e}{r}\right)^{l+1} \Delta H_{lma} \quad (3)$$

In this equation the coefficients k'_l are the load Love numbers, ρ_w and ρ_e are the mean density of sea water and the Earth. The spectral representation of the convolution operator that translates the mass layer effect in the geoid effect is shown in Fig. 3. From this figure it can be seen that the convolution procedure amplifies the lower degrees compared to the higher degrees. This mechanism allows to

identify whether a geophysical signal can be observed by GRACE as is discussed for instance in Wahr *et al.* (1998). Equally important is whether corrections that need to be applied to be able to observe these “science signals” can be applied with sufficient accuracy. In the following subsections we will try to answer this question by considering simulated errors of existing ocean tide models and atmospheric pressure models which are two large contributors in the GRACE data reduction scheme. (ie. modelling activities before the monthly geoid maps are provided to the scientific community).

4.1. SIGNAL CONTAMINATION

4.1.1. *Ocean Tides*

In order to compute the geoid effect caused by an ocean tide signal during the GRACE data reduction the Δh terms in Eq. (3) are replaced by recent tide models which are enhanced by TOPEX/Poseidon altimetry data, see also Fu and Cazenave (2001). To simulate a geoid model error as a result of still remaining ocean tide model errors we have assumed that the mass layer input function becomes the difference between the GOT99.2 and the FES99 model developed by Ray (1999) and Lefèvre *et al.* (2002) respectively. The amplitude maps of the self attraction geoid effect as a result of these model differences are shown in Fig. 4 for the tidal constituents M_2 , S_2 , O_1 and K_1 . It can be seen that the simulated geoid error in the open oceans is less than 0.5 mm. In polar regions and on continental shelf areas there are more significant differences that can reach the 3 mm level. These geoid errors are caused by the fact that the T/P inclination of 66 degrees limits the altimeter mapping range. Moreover in shallow water the altimeter track spacing is too coarse to map the finer details of the shallow water tides.

Attempts to design methods for assessing the tide model error contamination are discussed by Ray *et al.* (2002) and Knudsen and Andersen (2002) and Schrama (1995). The aliasing problem for a gravity mission is far more difficult to comprehend than the way tidal modelling errors map along repeating T/P altimeter ground tracks.

For GOCE gradiometry one could in principle use a similar technique as described in Schrama and Ray (1994) that relies on a harmonic analysis of repeat track altimetry. Unfortunately the gradiometer instrument noise level is large compared to the tide model signal, see also Schrama (1995). Noteworthy is that mapping of the S_2 tide from a sun-synchronous orbit like used for GOCE results in a situation where the corresponding geoid error is sampled at the same local solar time. The consequence is that an oscillating phenomenon results in a static field. Other tidal constituents result in shorter aliasing periods and can in principle be recovered by harmonic analysis, provided that there is sufficient mission science data. Within this context it should be mentioned that it took about 3 years of data to optimize the tide models from T/P altimeter data. To fully explain the tidal aliasing

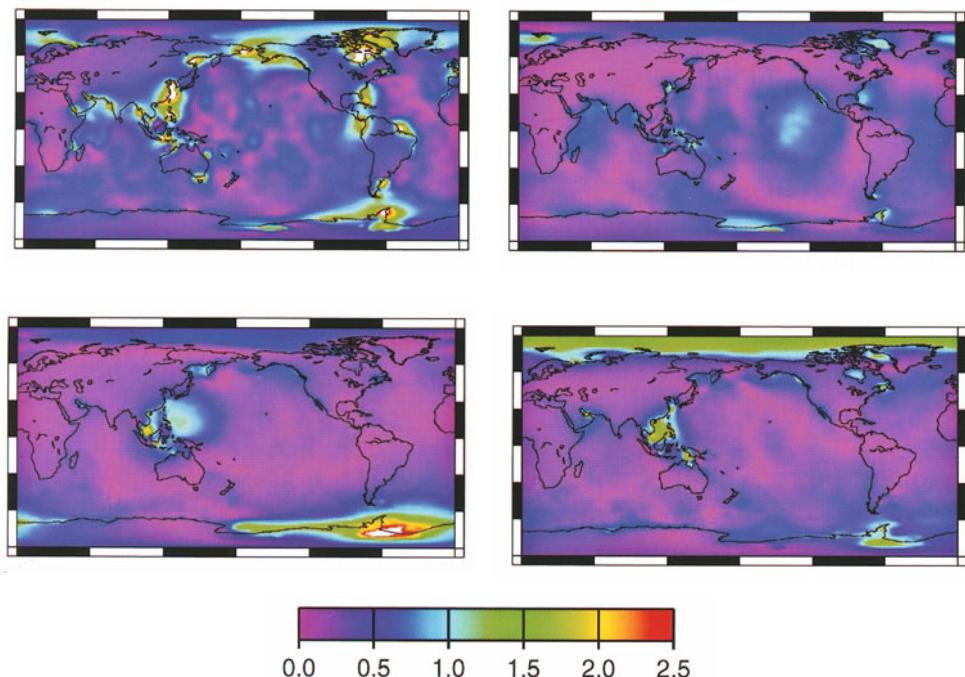


Figure 4. Self attraction error as a result of the difference between the ocean tide models GOT99.2 and FES99. The color scale goes from 0 to 2.5 mm. The order of the figures is M2: upper left, S2: upper right, O1: lower left, K1: lower right.

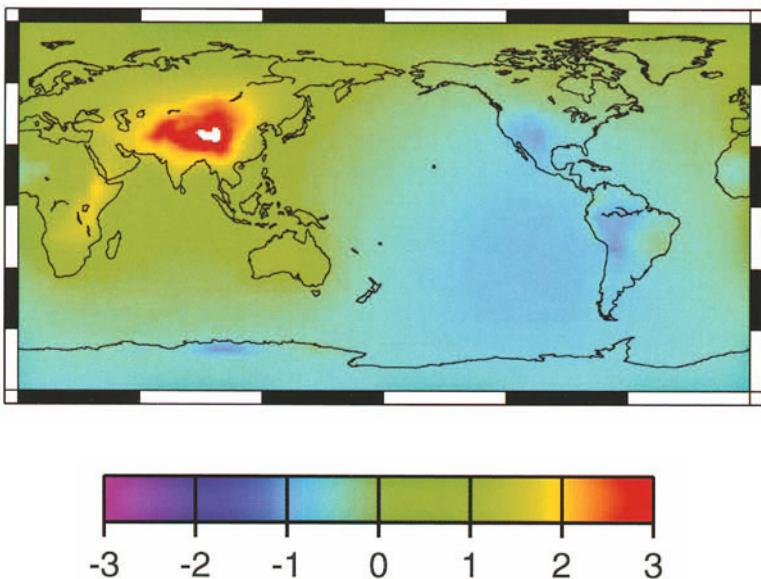


Figure 5. Mean geoid effect as the result of ECMWF vs NCEP pressure differences in the month of January 1992, the color scale runs from -3 to 3 mm.

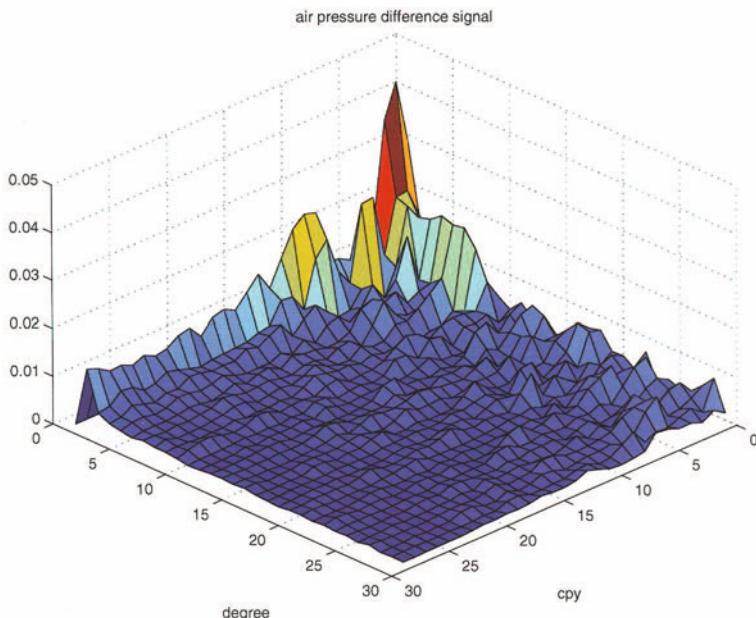


Figure 6. Wavenumber frequency difference spectrum of detrended air pressure in mbar for ECMWF vs NCEP over 1992, the left diagonal axis represent spherical harmonic degrees l , the right diagonal axis frequency in steps of cycles per year (cpy), the vertical axis represents mbar (formally mbar per square root of the spatial frequency and per square root of the temporal frequency).

problem for GOCE one should investigate the mapping of tidal model errors in the orbit perturbation analysis: for details see Schrama (1995).

For GRACE the tides are mapped in an even more complicated way since the orbital plane is not sun-synchronous but more optimized towards a polar orbit. Moreover tidal modeling errors map as velocity perturbations that eventually end up in the GRACE data reduction scheme. According to Ray *et al.* (2002) an important factor is the rate of change of the orbital plane of GRACE compared to the rate of change of the tide generating potential at a specified constituent. The tidal aliasing periods mentioned in Ray *et al.* (2002) are that K_1 maps at 7.48 years while S_2 maps at 161 days.

It is the opinion of the author that tidal aliasing justifies further research on the propagation of the simulated tide model errors in the adopted data reduction procedures implemented for GRACE. Similar activities are required for GOCE once a decision has been made on the final data reduction method.

4.1.2. Air Pressure

The motivation for studying the self attraction geoid effect that follows from the air pressure signal is provided in Fig. 2. In this case the thick line is representing the total contribution of the air pressure correction of the NCEP (National Centers for Environmental Prediction) reanalysis data in 1992. Interestingly enough, the

magnitude of this correction is up to a factor 100 larger than the best GRACE curve and the natural question arises whether air pressure can be modelled with sufficient accuracy to fully exploit the GRACE sensitivity at lower degrees.

The validity of the atmospheric pressure correction algorithm is addressed in Velicogna *et al.* (2001) where it is stated that the pointwise accuracy of global models is of the order of 1.0 to 1.5 mbar. They conclude this from a comparison of ECMWF data (European Center for Medium Range Weather Forecasts), NCEP reanalysis data and in-situ air pressure data. In Velicogna *et al.* (2001) it is also suggested that the accuracy of surface pressure corrections may be improved after spatial and temporal smoothing of the input data. They show examples where the air pressure error is reduced to about 0.3 to 0.5 mbar (in remote deserts) when the average field is computed over a period of a month. The conclusion of Velicogna *et al.* (2001) is that regional improvement by in-situ air pressure measurements appears to be sufficient to remove the air pressure effect from GRACE data.

In the following we simulate the air pressure model errors by the difference between the ECMWF and the NCEP reanalysis sea level pressure data provided as daily grids in 1992. In Fig. 5 we show the mean geoid error effect as a result of this model error; in this computation the sea level pressure changes are converted to equivalent water height values over land assuming a vertical air pressure gradient by an exponential decay law and 100% inverse barometric compensation over the oceans. Geoid error grids are then computed on a daily basis and averaged over a period of 12 months. The polar regions beyond 70 degrees latitudes were deliberately left out of this analysis since it is assumed that the discrepancy between the ECMWF and NCEP models is unrealistic, see also Verhagen (2001). This computation shows that the error is non-uniform and that it is mostly contained in the Himalayas, in Africa and to a lesser extent in Australia. It is for this reason that we expect that air pressure contamination is significant and that it may show up as a systematic bias in a static gravity model.

A wavenumber frequency analysis of the simulated air pressure differences minus the average effect is shown in Fig. 6. This spectrum represents on one axis the spherical harmonic degree and on the other axis the frequency in cycles per year (cpy). It shows that systematic long wavelength spatial and temporal residuals between both meteorologic models exist. An integration in the wavenumber frequency spectrum learns that the signal contained at frequencies longer than two months (i.e. twice the monthly mapping cycle by GRACE) and over spherical harmonic degrees up to 20 results in an air pressure rms of 0.16 mbar. The magnitude of this modelling error appears to be significant in view of the anticipated accuracy of GRACE which promises a geoid to be mapped to within 0.03 mm below degree and order 20. Furthermore it should be mentioned that air pressure changes will occur within the GRACE mapping cycle. Also these signals will alias (or fold) into the monthly geoid maps. So far integration in the wavenumber frequency spectrum suggests that such effects are far smaller than the above mentioned effects. This

analysis does not account for atmospheric tides, see also Ray *et al.* (2002), due to the daily sampling rate of our input maps.

5. Conclusions

In this paper we focus on the problem of the error characteristics of gravity fields from CHAMP, GRACE and GOCE including the EGM-96 solution in order to be able to optimize the use of this information in the proposed scientific experiments. The paper is written in three parts from which the main conclusions are summarized below.

The gravity solution of the EGM-96 model and the CHAMP solution both come with their formal error covariance matrices. The GRACE and GOCE mission performances are currently based upon a covariance propagation technique and are usually displayed in the form of degree variances. A potential drawback of a representation in degree variances is that it assumes an isotropic and homogeneous covariance function on the sphere. The geographical representation of geoid errors from the EGM-96 solution indicates for instance that the distribution of errors is correlated to the quality and coverage of the input data.

A similar problem occurs in the combination of gravity normal matrices such as the EIGEN-1S solution and existing normal matrices; the EIGEN-1S gravity model is tailored to the CHAMP satellite. Optimal data combination strategies such as variance component estimation techniques that properly balance existing gravity normal matrices relative to the new normal matrix appear to offer a solution for this problem.

An important application of accurate gravity models that has made its way into the GOCE acronym is to improve our knowledge of the mean oceanic circulation field. Research by LeGrand (2001) and others suggest that GOCE data leads to a 30% to 60% reduction in mean oceanic transport errors. Nevertheless it is currently also possible to assimilate only the long wavelength information from satellite altimetry and geoid reference models in hydrographic inverse models. A careful setup of the cost function in the assimilation algorithm by geoid errors that follow from the EGM-96 covariance matrix have been presented in LeGrand *et al.* (2002). Further research is required to investigate whether eddy resolving ocean models can be applied to improve the location of the mean climatological path of the Gulf stream or whether straightforward averaging is adequate to reduce the effect of oceanic mesoscale variability in estimating the steady state ocean circulation.

A second series of applications concern the observation of temporal geoid effects caused by geophysical processes such as changes in the continental water balance. We discuss the properties of the kernel that describes the relation between a surface mass layer and the geoid to conclude that most of the signal can be expected up to degree and order 30 in the gravity field.

The scientific interest in GRACE is in the recovery of gravity signals such as variations in the continental water balance. To accomplish this task reductions should be made for the variations caused by tides and air pressure variations. In order to quantify this problem we discuss the results of simulated errors as a result of tide model differences and air pressure model differences. Our conclusion is that tidal modelling errors occur with a magnitude up to 3 mm where the K_1 constituent is likely to map at a frequency that exceeds the planned 5 year length of the GRACE mission. From the air pressure error simulation we conclude that tides appear to be more significant than the errors introduced by the air pressure correction algorithm. The expected signal error in the simulated air pressure signal is estimated at 0.16 mbar for periods longer than 2 months and for spherical harmonic degrees less than 20. Such pressure errors appear to be significant in view of the anticipated GRACE sensitivity of 0.03 mm in the geoid below degree 20.

Both issues justify future research with regard to the implementation of algorithms for separation of signal and noise from monthly GRACE solutions.

Acknowledgements

I thank Pieter Visser, Pascal le Grand, Johannes Bouman, Nico Sneeuw and Richard Ray for their helpful comments and/or data used in this paper. The ECMWF air pressure grids originate from Meteo France provided to TUD/DEOS through a cooperation with CNES/GRGS, Cathy Smith at NOAA/CDC assisted in getting access to their reanalysis data.

References

- Colombo, O.L.: 1986, *The global mapping of gravity with two satellites*, Netherlands Geodetic Commission, New Series No. 30, pp. 1–180.
- Dickey, J.O.: 1997, *Satellite Gravity and the Geosphere*, National Research Council Report, Nat. Acad. Washington D.C., 112 pp.
- ESA: 1999, *Gravity field and steady state ocean circulation mission*, ESA SP-1233, pp 217.
- Fu, L.L., et al.: 1994, ‘TOPEX/POSEIDON mission overview’, *JGR* **99**, 24,369–24,382.
- Fu, L.L. and Cazenave, A.: 2001, *Satellite altimetry and Earth Sciences, A handbook of techniques and applications*, International Geophysics Series, Vol 69, Academic Press.
- Ganachaud, A., Wunsch, C., Kim, M.-C., and Tapley, B.: 1997, ‘Combination of TOPEX/POSEIDON data with a hydrographic inversion for determination of the oceanic general circulation and its relation to geoid accuracy’, *Geophysical J. Int.* **128**, 708–722.
- Heiskanen, W.A. and Moritz, H.: 1979, *Physical Geodesy*, Reprint Institute of Physical Geodesy, TU Graz, Austria.
- Hurlburt, H.E. and Thompson, J.D.: 1980, ‘A numerical study of Loop Current intrusions and eddy-shedding’, *J. Phys. Oceanogr.* **10**, 1611–1651.
- Knudsen, P. and Andersen, O.: 2002, ‘Correcting GRACE gravity fields for ocean tide effects’, *Geophysical Research Letters* **29**, 19.1–19.4.

- Koch, K.-R.: 1986, 'Maximum likelihood estimate of variance components', *Bulletin Geodesique* **60**, 329–338.
- Koch, K.-R.: 1990, *Bayesian Inference with Geodetic Applications*, Lecture notes in Earth Sciences, Springer, Berlin.
- Lefèvre, F., Lyard, F., Le Provost, C., and Schrama, E.: 2002, 'FES99: A global tide finite element solution assimilating tide gauge and altimetric information', *Journal of Atmospheric and Oceanic Technology* **19**, 1345–1356.
- Lerch, F.J.: 1991, 'Optimum data weighting and error calibration for estimation of gravitational parameters', *Bulletin Geodesique* **65**, 44–52.
- LeGrand, P.: 2001, 'Impact of the Gravity Field and Steady State Ocean Circulation Explorer (GOCE) mission on ocean circulation estimates: Volume fluxes in a climatological inverse model of the Atlantic', *Journal of Geophysical Research* **106**, 19,597–19,610.
- LeGrand, P., Schrama, E., and Tournadre, J.: 2002, 'An Inverse Estimate of the Dynamic Topography of the Ocean', *Geophysical Research Letters*, 30(2), 34-1, cite ID 1062, DOI 10.1029/2002GL014917
- Lemoine, *et al.*: 1998, 'The development of the joint NASA GSFC and NIMA geopotential model EGM-96 gravity model', NASA/TP-1998-206861.
- Lillibridge, J., *et al.*: 2000, 'ERS-2 altimetry in operational NOAA forecast models', Proc. 4th ERS Symp., Gothenberg, Sweden.
- Moritz, H.: 1980, *Advanced Physical Geodesy*, H. Wichman Verlag, Karlsruhe.
- NRL: 2002, <http://www.ocean.nrlssc.navy.mil/altimetry/>.
- Ray, R.: 1999, 'A Global Ocean Tidel Model from TOPEX/POSEIDON Altimetry: GOT99.2', NASA TM 1999-209478.
- Ray, R., Rowlands, D., and Egbert, G.: 2003, 'Tidal models in a New Era of Satellite Gravimetry', *Space Science Reviews*, this volume.
- Reigber, *et al.*: 2002, 'A High-Quality Global Gravity Field Model from CHAMP GPS Tracking Data and Accelerometry (EIGEN-1S)', *Geophysical Research Letters* **29**, 10129/2002GL015064.
- Schrama, E.: 1991, 'Gravity Field Error Analysis: Applications of Global Positioning System Receivers and Gradiometers on Low Orbiting Platforms', *Journal of Geophysical Research* **96**, 20,041–20,051.
- Schrama, E. and Ray, R.: 1994, 'A preliminary tidal analysis of TOPEX/POSEIDON altimetry', *JGR* **99**, 24,799–24,808.
- Schrama, E.: 1995, 'Gravity missions reviewed in the light of the indirect ocean tide potential', Proceedings IAG symposia proceedings No 116, 131–140, Springer Verlag.
- Schröter J., Losch, M., and Sloyan, B.: 2002, 'Impact of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission on ocean circulation estimates 2: Volume and heat fluxes across hydrographic section of unequally spaced stations', *JGR* **107**, 4.1–4.20.
- Velicogna, I., Wahr, J., and van den Dool, H.: 2001, 'Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals?', *JGR* **106**, 16,415.
- Verhagen, S.: 2001, *Time variations in the gravity field, the effect of the atmosphere*, MSc report, 81 pages, TU Delft Geodesy.
- Wahr, J., Molenaar, M., Bryan, F.: 1998, 'Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE', *JGR* **103**, 30,205–30,229.
- Wunsch, C.: 1996, *The Ocean Circulation Inverse Problem*, Cambridge University Press.

ESTIMATING THE HIGH-RESOLUTION MEAN SEA-SURFACE VELOCITY FIELD BY COMBINED USE OF ALTIMETER AND DRIFTER DATA FOR GEOID MODEL IMPROVEMENT

SHIRO IMAWAKI (imawaki@riam.kyushu-u.ac.jp)

Research Institute for Applied Mechanics, Kyushu University, Japan

HIROSHI UCHIDA (huchida@jamstec.go.jp)

Japan Marine Science and Technology Center, Japan

KAORU ICHIKAWA (ichikawa@riam.kyushu-u.ac.jp)

Research Institute for Applied Mechanics, Kyushu University, Japan

DAISUKE AMBE (ambe@riam.kyushu-u.ac.jp)

Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Japan

Received: 7 June 2002; Accepted in final form: 26 September 2002

Abstract. The mean sea-surface height obtained from satellite altimeters is different from the geoid by the amount of mean sea-surface dynamic topography associated with ocean currents. Assuming geostrophy at the sea surface, the mean sea-surface dynamic topography can be obtained from the mean sea-surface velocity field. This field is derived by combining anomalies (i.e., deviations from the mean) of sea-surface velocity obtained from altimeter data and *in situ* surface velocities estimated from trajectories of surface drifting-buoys (hereafter, drifters). Where a drifter measured the surface velocity, the temporal mean velocity can be estimated by subtracting the altimeter-derived velocity anomaly at that time from the drifter-measured surface velocity. The method is applied to the surface flow field of the North Pacific, using TOPEX/POSEIDON and ERS-1/2 altimeter data, and WOCE-TOGA surface drifter data obtained from October 1992 through December 2000. The temporal mean velocity field is estimated with a resolution of quarter degrees in both latitude and longitude. The obtained mean velocity field clearly shows the Kuroshio and Kuroshio Extension, which are narrower and stronger than the climatological mean features derived from historical hydrographic data averaged over several decades. Instantaneous velocities are estimated by summing up these temporal mean velocities and anomalies, every ten days during the eight years. They compare well with *in situ* velocities measured by the surface drifters. The instantaneous velocity field shows energetic fluctuation of the Kuroshio Extension vividly.

1. Introduction

Recently, the sea-surface topography has been observed very precisely by satellite altimeters, and fairly precise geoid has been obtained. The mean sea-surface height obtained from satellite altimeters, however, is different from the geoid by the amount of sea-surface dynamic topography associated with ocean currents. Therefore, the precise mean sea-surface dynamic topography should be obtained in order to estimate the geoid precisely.



Space Science Reviews **108:** 195–204, 2003.

© 2003 Kluwer Academic Publishers.

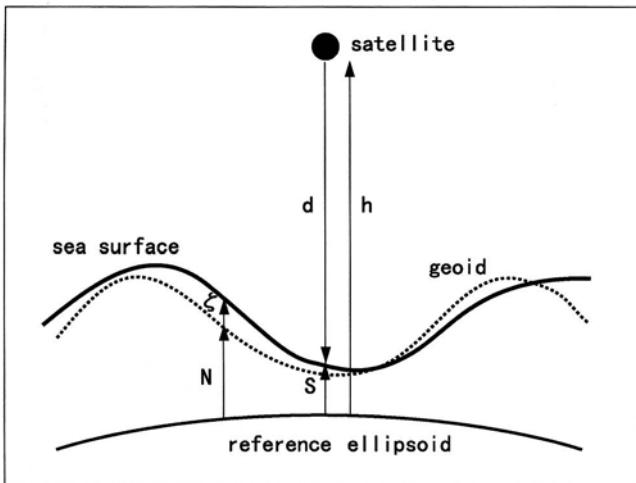


Figure 1. Conceptual view of satellite altimeter measurements.

Here we try to detect the undulation of the mean sea-surface dynamic topography with high resolution, down to the oceanic mesoscale. Assuming geostrophy at the sea surface, the high-resolution mean sea-surface dynamic topography can be obtained from the high-resolution mean sea-surface velocity field. This field is derived by combining anomalies (i.e., deviations from the mean) of sea-surface velocity obtained from altimeter data and *in situ* surface velocities estimated from trajectories of surface drifting-buoys (hereafter, drifters). The present method to estimate the mean velocity field from the combined use of altimeter and drifter data has recently been developed by Uchida and Imawaki (2003) on the basis of Uchida et al. (1998).

2. Method

The satellite altimeter measures the distance d from the spacecraft to the sea surface. The height h of the spacecraft from the reference ellipsoid is measured by a satellite tracking system. Those two properties give the sea-surface height S_e from reference ellipsoid (Fig. 1). The sea-surface height consists of the geoid height N and sea-surface dynamic topography ξ . Namely,

$$S_e = N + \xi + \epsilon \quad (1)$$

where ϵ is the measurement error. If the measurement is repeated at a given interval, the observed quantity ϕ can be divided into the temporal mean $\bar{\phi}$ and anomaly ϕ' (or the deviation from the mean); $\phi = \bar{\phi} + \phi'$. For the sea-surface dynamic topography,

$$\xi = \bar{\xi} + \xi' \quad (2)$$

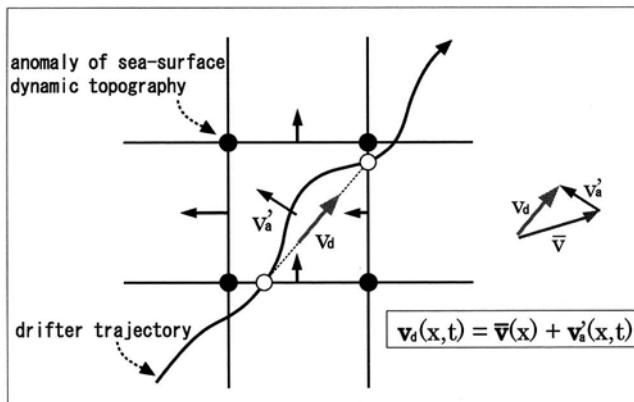


Figure 2. Conceptual view of estimating the mean velocity $\bar{V}(x)$ by combining altimeter-derived velocity anomaly $V'_a(x, t)$ and drifter-derived instantaneous velocity $V_d(x, t)$ at a given grid box. The altimeter-derived velocity anomaly is interpolated at the time when the drifter passed through the grid.

Using these, the mean \bar{S}_e and anomaly S'_e of the measured sea-surface height can be expressed as follows;

$$\bar{S}_e = N + \bar{\zeta} + \bar{\epsilon} \quad (3)$$

$$S'_e = \zeta' + \epsilon' \quad (4)$$

where the geoid height is assumed to be constant in time. Therefore, you need the accurate geoid height when you estimate absolute sea-surface dynamic topography ζ using Equation (1) or mean sea-surface dynamic topography $\bar{\zeta}$ using Equation (3), from measured sea-surface height. On the other hand, you do not need geoid information when you estimate only anomaly ζ' of sea-surface dynamic topography from measured sea-surface height, as shown in Equation (4); it can be estimated with an error similar to the altimeter measurement error.

Figure 2 schematically shows how the temporal mean velocity can be estimated by combined use of the altimeter data and intermittent surface velocity data. From gridded data of anomalies ζ' of sea-surface dynamic topography, you can easily estimate the anomaly $V'_a = (u', v')$ of sea-surface geostrophic velocity as follows;

$$u' = -(g/f)\zeta'_y \quad (5)$$

$$v' = (g/f)\zeta'_x \quad (6)$$

where g is the gravity of the earth, f the Coriolis parameter expressing the effect of rotation of the earth, ζ'_x the zonal derivative of anomaly ζ' of sea-surface dynamic topography, and ζ'_y the meridional one.

If you measure the *in situ* sea-surface velocity V_d at some point x at some time t , you can estimate the mean sea-surface velocity \bar{V} at that point by combining anomaly V'_a of sea-surface geostrophic velocity at that point at that time, as follows;

$$V_d(x, t) = \bar{V}(x) + V'_a(x, t) \quad (7)$$

From this mean velocity field, you can easily estimate the mean sea-surface dynamic topography $\bar{\zeta}_e$ assuming again geostrophy at the sea surface.

Finally, you can estimate the geoid height N from this estimated mean sea-surface dynamic topography $\bar{\zeta}_e$ and the mean \bar{S}_e of the measured sea-surface height, using Equation (3), as follows;

$$N = \bar{S}_e - \bar{\zeta}_e + \tilde{\epsilon} \quad (8)$$

where $\tilde{\epsilon}$ is the estimated error of $\bar{S}_e - \bar{\zeta}_e$. In most of previous studies, the measured mean sea-surface height \bar{S}_e is assumed to be equal to the geoid height N , ignoring the mean sea-surface dynamic topography $\bar{\zeta}_e$. The latter must be taken into account, especially when the high-resolution geoid is concerned.

3. Data

We use satellite altimeter data from TOPEX/POSEIDON and the European Remote Sensing satellite (ERS)-1/2. TOPEX/POSEIDON has been collecting the sea-surface height data from September 1992, with a repeat period of 9.9 days. The ERS-1/2 have been collecting the altimeter data from April 1992, with a repeat period of 35 days; during some periods, the repeat cycles were different. Here we use the map of sea-surface height anomaly S'_e prepared by the Collect, Localization, Satellites (CLS) in France from October 1992 to December 2000 (AVISO, 1997); the sea-surface height anomalies obtained by TOPEX/POSEIDON and ERS-1/2 were merged and gridded with a resolution of 0.25° latitude \times 0.25° longitude globally every ten days, after removal of high frequency fluctuations including tides (Le Traon et al., 1998). Here an additional error has been introduced to the gridded data S'_e because time-varying sea-surface height anomalies observed during 10 days are folded and averaged to provide a single snapshot representing that 10-day period. The anomaly $V'_a(x, t) = (u', v')$ of sea-surface geostrophic velocity is estimated from the anomaly ζ' of sea-surface dynamic topography, which is practically equivalent to the sea-surface height anomaly S'_e , using Equations (5) and (6), as shown schematically in Fig. 2.

We use trajectories of surface drifters to estimate surface velocities. Figure 3 shows all the trajectories of drifters (about 1,600 for the North Pacific) used in this study; they were obtained by the Surface Velocity Program of the World Ocean Circulation Experiment (WOCE) and Tropical Ocean and Global Atmosphere (TOGA) during the same period as the above-mentioned analysis period of about eight years. Drifters were equipped with drogues centered at 15 m depth and tracked by

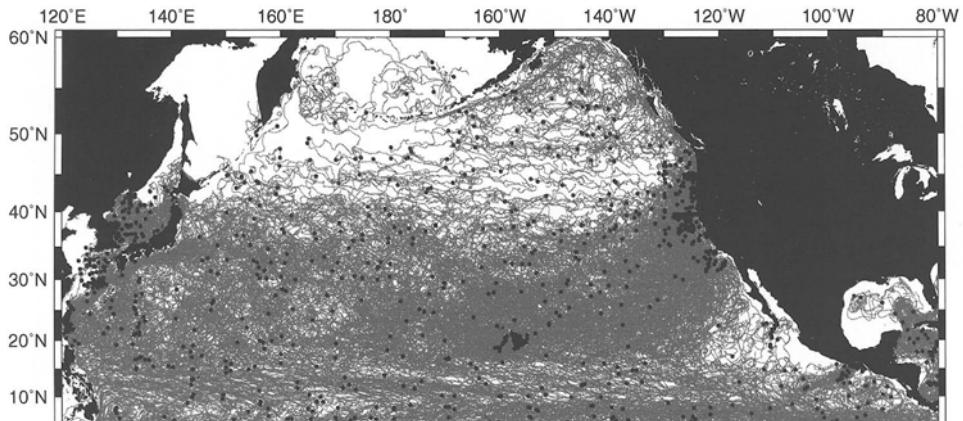


Figure 3. Trajectories of all the drifters (about 1,600) during 1992–2000 used in the present study. Dots are starting points of individual trajectories.

satellites. The trajectories were smoothed by a low-pass filter and sub-sampled every 6 hours (Hansen and Poulain, 1996). The surface velocity $V_d(x, t)$ is estimated at the box of 0.25° latitude $\times 0.25^\circ$ longitude through which a drifter passed, as shown in Fig. 2. The Ekman flow in this sub-surface layer at that time is estimated by using a method similar to Lagerloef et al. (1999) with ERS and Advanced Earth Observing Satellite (ADEOS) scatterometer wind data (Anonymous, 2001), and subtracted from the estimated drifter velocity.

4. Mean Flow Field

The mean surface velocity field during seven years of 1993–1999 for the Kuroshio and Kuroshio Extension regions is shown in Fig. 4. Here only the mean values which are averaged over three or more estimates of mean velocity $\bar{V}(x)$ at individual quarter degree grid points are used; more than twenty estimates of \bar{V} are obtained in the Kuroshio region south of Japan. Errors in speeds of mean velocities are estimated at individual grid points. Their average over the entire North Pacific is 0.05 m/s, which is satisfactorily small, compared with the speed of the mean velocity itself (0.16 m/s) averaged over the same area (Uchida and Imawaki, 2002).

The flow field precisely shows the major features of this strong western boundary current system, which have been documented fragmentally from place to place. The mean width of the strong current is about 100 km. There are several anti-cyclonic gyres on the offshore side of the Kuroshio. The Kuroshio Extension takes the stationary meander pattern, when leaving the Japanese coast. The mean Kuroshio and Kuroshio Extension are narrower and stronger than the climatological mean Kuroshio and Kuroshio Extension derived from historical hydrographic data averaged over several decades (Hasunuma and Yoshida, 1978; Boyer and Levitus,

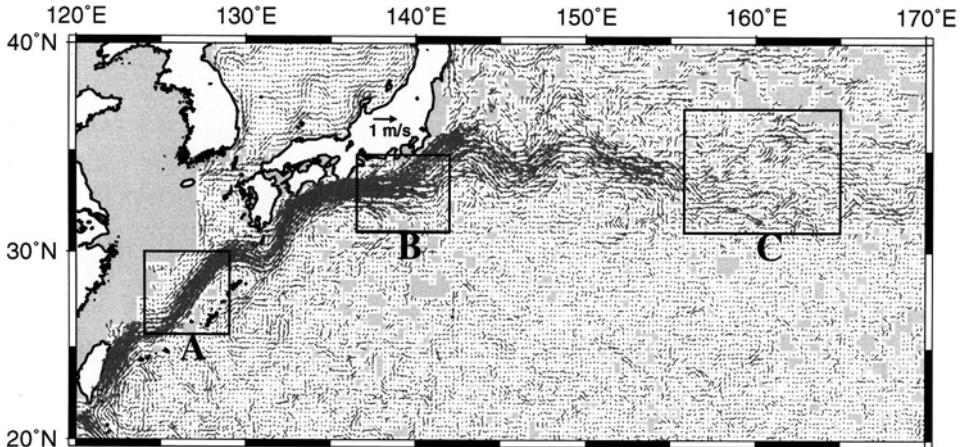


Figure 4. Mean (1993–1999) sea-surface velocity field of the Kuroshio and Kuroshio Extension regions estimated from the combined use of altimeter and drifter data. The vector scale is shown by the arrow in the upper middle. Shade shows grids where no reliable means are available because of the lack of drifter data (except for the East China Sea). See text for boxes A, B and C.

1997). The details of the advantage of combining the altimeter data and drifter data are given by Uchida and Imawaki (2003). The pronounced weakness of the velocity of the Kuroshio Extension, compared with the Kuroshio south of Japan, will be discussed in the next section.

5. Instantaneous Flow Field

Once you obtain the temporal mean velocity $\bar{V}(x)$, you can easily obtain the instantaneous surface geostrophic velocity V_g at point x at time t , from the anomaly $V'_a(x, t) = (u', v')$ of sea-surface geostrophic velocity, as follows;

$$V_g(x, t) = \bar{V}(x) + V'_a(x, t) \quad (9)$$

Those instantaneous velocity fields are obtained every ten days from October 1992. As an example, the anomaly field of sea-surface geostrophic velocity for the Kuroshio and Kuroshio Extension regions on May 9, 1993 is shown in Fig. 5; the flow field shows an average field over ten days centered on that day. The instantaneous velocity field on the same day is shown in Fig. 6. The instantaneous velocity field clearly shows the propagating meander of the Kuroshio south of Japan, and energetic fluctuations in the Kuroshio Extension region with meanders and rings. Note that the drifter-derived velocity data are sparse during the individual 10-day intervals.

The boxes labeled A, B and C in Figs. 4, 5 and 6 show three typical areas where contributions of the mean and anomaly fields to the instantaneous velocity field are different. In box A, the instantaneous field is determined mostly by the mean field;

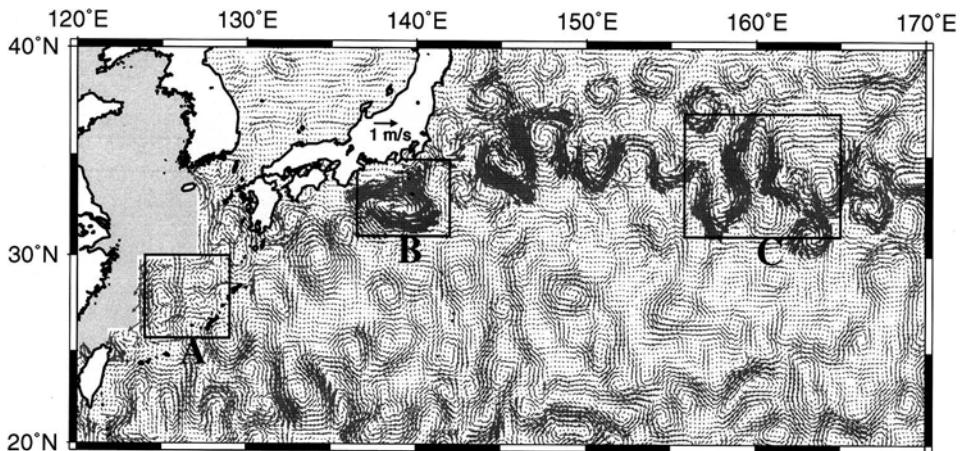


Figure 5. Same as Fig. 4 but for the anomaly field on May 9, 1993 estimated from the altimetry data.

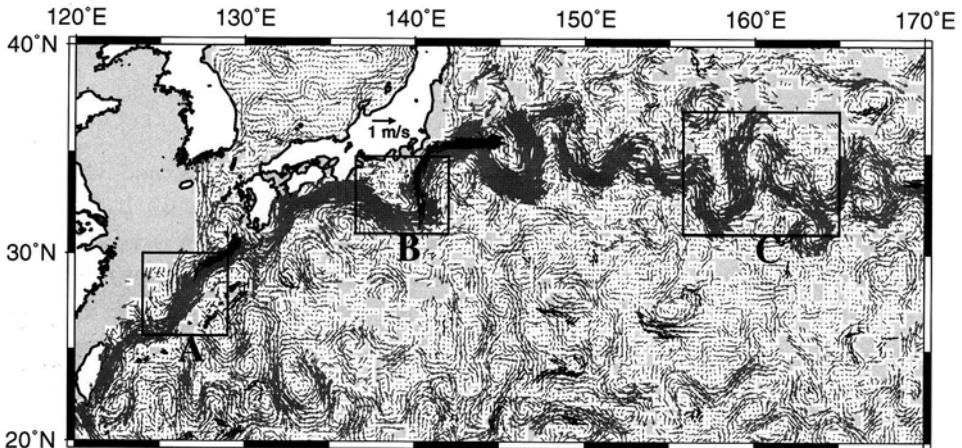


Figure 6. Same as Fig. 5 but for the instantaneous velocity field obtained as the sum of Figs. 4 and 5. Black arrows are drifter-derived velocity vectors obtained from 37 drifters during the 10 days centered on that day.

there, the mean velocity field must be estimated accurately. On the other hand, in box C, it is determined mostly by the anomaly field; there, the mean field may not necessarily be estimated accurately. In box B, both the mean and anomaly fields are important.

The instantaneous velocity field obtained every ten days during the eight-year period of 1993–2000 shows energetic fluctuation of the Kuroshio Extension vividly. An interesting feature is that the mean Kuroshio Extension almost disappears just before the Shatsky Rise located at about 155–160°E, although the instantaneous Kuroshio Extension continues to flow down to the Emperor Seamounts located at about 170°E. Animation of the surface flow field can show those circumstances clearly.

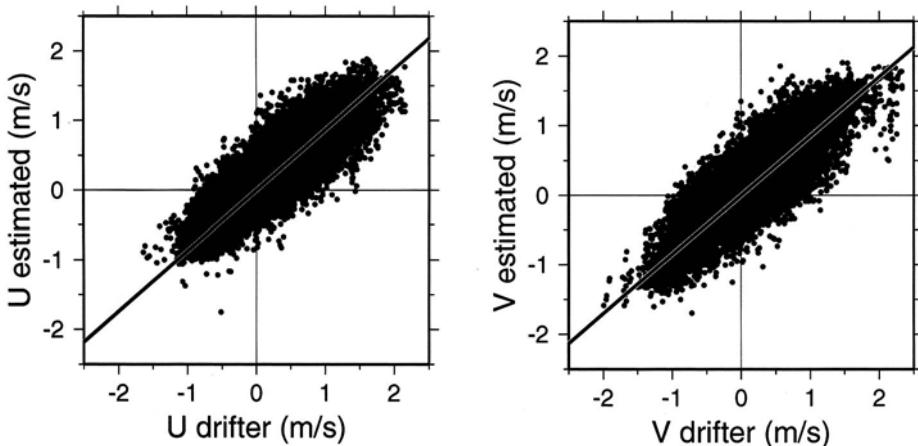


Figure 7. Comparison of zonal u and meridional v components of the estimated instantaneous velocity (ordinate) with those of the drifter velocity (abscissa). The number of data is 440,809.

The velocity of the mean Kuroshio Extension between 140°E and 150°E is weaker than that of the mean Kuroshio south of Japan (see Fig. 4). It does not necessarily mean, however, that the flow of the Kuroshio Extension is weaker than that of the Kuroshio south of Japan. It is merely because the location of the Kuroshio Extension fluctuates meridionally and so the long-term mean of the velocity is smaller than the instantaneous velocity, while the location of the Kuroshio south of Japan is more stable.

The mean flow field shown in Fig. 4 is somewhat artificial, and is difficult to be compared with velocities observed by other methods. On the other hand, the instantaneous flow field can be compared with other data easily. Here we compare the zonal u and meridional v components of the estimated instantaneous velocity $V_g(x, t)$ with those of the drifter velocity $V_d(x, t)$. The result is shown in Fig. 7. The comparison is satisfactory; the slope of regression line is close to unity (0.87 for u and 0.85 for v), the root-mean-square (rms) difference from the line is small (0.11 m/s for both), compared with the dynamic range of about 3 m/s, and the correlation coefficient is high (0.87 for u and 0.79 for v). The fact that the slope of regression line is smaller than unity is probably due to smoothing of the altimeter data, which is inevitable in constructing gridded data by an interpolation. The rms difference of 0.11 m/s includes both the error of the estimated mean velocity and difference of anomalies between the altimeter-derived velocity and drifter-measured velocity. It is satisfactorily small, compared with the typical speed of the Kuroshio (1 to 2 m/s) and velocity of strong mesoscale eddies. Note, however, that those two properties are not independent with each other, because the drifter velocity has been used to estimate the instantaneous velocity.

6. Summary

Here is shown how the geoid height can be improved by estimating the mean sea-surface dynamic topography as accurate as possible. Assuming geostrophy at the sea surface, the high-resolution mean sea-surface dynamic topography can be obtained from the high-resolution mean sea-surface velocity field. This field is derived by combining anomalies of sea-surface velocity obtained from altimeter data and *in situ* surface velocities estimated from trajectories of surface drifters. Where a drifter measured the surface velocity, the temporal mean velocity is estimated by subtracting the altimeter-derived velocity anomaly at that time from the drifter-measured surface velocity.

The method is applied to the sea-surface flow field of the North Pacific, using TOPEX/POSEIDON and ERS-1/2 altimeter data, and WOCE-TOGA surface drifter data obtained from October 1992 through December 2000. The temporal mean velocity field is estimated with a resolution of quarter degrees in both latitude and longitude. The obtained mean velocity field clearly shows the Kuroshio and Kuroshio Extension, which are narrower and stronger than the climatological mean Kuroshio and Kuroshio Extension derived from historical hydrographic data averaged over several decades.

The instantaneous velocities are estimated by summing up these temporal mean velocities and anomalies, every ten days during the eight years. They compare well with velocities measured by the surface drifters. The instantaneous velocity field shows energetic fluctuation of the Kuroshio Extension vividly.

In future, the accuracy of the estimated mean field should be examined in more detail, and the mean sea-surface dynamic topography should be estimated from the velocity field assuming geostrophy. The present method can be applied all over the world except for the Equatorial regions (between about 5°N and 5°S), where the approximation of geostrophy is not valid.

When we have a high-resolution mean sea-surface dynamic topography $\bar{\zeta}_e$ using the present method, we can estimate a precise geoid height N from a mean sea-surface height \bar{S}_e measured by altimeters, using Equation (8). Note, however, that the accurate mean sea-surface height \bar{S}_e is obtained only along subsatellite tracks of the altimeters and therefore the resolution of the geoid is substantially limited by the distribution of those subsatellite tracks. The corrected geoid height mentioned above will be a data set with which we can compare new precise geoids which will be obtained soon by satellites measuring the gravity and hence the geoid itself, instead of the sea-surface height.

Acknowledgements

We used the map of sea-surface height anomaly prepared by the Collect, Localization, Satellites (CLS), and distributed by the Archiving, Validation and Inter-

pretation of Satellites Oceanographic Data (AVISO) in France. We used surface drifter data prepared by the Atlantic Oceanographic and Meteorological Laboratory of National Oceanic and Atmospheric Administration in U.S.A., and distributed by Marine Environmental Data Service (MEDS) in Canada. We thank Dr. Mark Drinkwater of the European Space Agency for his useful comments, which improved the manuscript. Ms. Miki Tsuruzono helped us to prepare the final formatted manuscript. This work was supported partly by the Core Research for Evolutional Science and Technology sponsored by the Japan Science and Technology Corporation.

References

- Anonymous: 2001, *Mean surface wind fields from the ERS-AMI and ADEOS-NSCAT microwave scatterometers, August 1991 to May 2000. WOCE Global Data, Version 2.0. CD-ROM*. WOCE International Project Office, Southampton, UK.
- AVISO: 1997, *AVISO Handbook: Sea Level Anomaly Files*, AVINT-011-312-CN, 21st ed. Rep., 24 pp.
- Boyer, T. P. and S. Levitus: 1997, 'Objective analyses of temperature and salinity for the world ocean on a 1/4° grid. *NOAA Atlas NESDIS 11*', U. S. Government Printing Office, Washington D.C.
- Hansen, D. V. and P.-M. Poulain: 1996, 'Quality control and interpolations of WOCE-TOGA drifter data'. *J. Atmos. Ocean. Tech.* **13**, 900–909.
- Hasunuma, K. and K. Yoshida: 1978, 'Splitting of the subtropical gyre in the western North Pacific'. *J. Oceanogr. Soc. Japan* **34**, 160–172.
- Lagerloef, G. S. E., G. T. Mitchum, R. B. Lukas, and P. P. Niiler: 1999, 'Tropical Pacific near-surface currents estimated from altimeter, wind and drifter data'. *J. Geophys. Res.* **104**, 23313–23326.
- Le Traon, P. Y., F. Nadal and N. Ducet: 1998, 'An improved mapping method of multisatellite altimeter data'. *J. Atmos. Ocean. Tech.* **15**, 522–534.
- Uchida, H. and S. Imawaki: 2003, 'Eulerian mean surface velocity field derived by combining drifter and satellite altimeter data'. *Geophys. Res. Lett.*, **30**(5), 1229, DOI: 10.1029/2002GLO16445.
- Uchida, H., S. Imawaki and J.-H. Hu: 1998, 'Comparison of Kuroshio surface velocities derived from satellite altimeter and drifting buoy data'. *J. Oceanogr.* **54**, 115–122.

COMBINED USE OF ALTIMETRY AND IN SITU GRAVITY DATA FOR COASTAL DYNAMICS STUDIES

K. HAINES¹, R. HIPKIN², C. BEGGAN³, R. BINGLEY⁴, F. HERNANDEZ⁵,
J. HOLT⁶, T. BAKER⁷ and R.J. BINGHAM¹

¹*Environmental Systems Science Centre, University of Reading, United Kingdom*, ²*Department of Geology & Geophysics, University of Edinburgh, United Kingdom*, ³*Formerly at Department of Meteorology, University of Edinburgh, United Kingdom*, ⁴*Institute of Engineering Geology & Space Geodesy, University of Nottingham, United Kingdom*, ⁵*Collecte Localisation Satellites, Ramonville St. Agne, France*, ⁶*Proudman Oceanographic Laboratory, Bidston Observatory, United Kingdom*
(kh@mail.nerc-essc.ac.uk)

Received: 7 June 2002; Accepted in final form: 30 September 2002

Abstract. Accurate local geoids derived from *in situ* gravity data will be valuable in the validation of GOCE results. In addition it will be a challenge to use GOCE data in an optimal way, in combination with *in situ* gravity, to produce better local geoid solutions. This paper discusses the derivation of a new geoid over the NW European shelf, and its comparison with both tide gauge and altimetric sea level data, and with data from ocean models. It is hoped that over the next few years local geoid methods such as these can be extended to cover larger areas and to incorporate both *in situ* and satellite measured gravity data.

1. Introduction

Several factors make the NW European shelf an ideal location to develop and test a marine geoid. There is a wealth of *in situ* gravity measurements which are available, much of which has been collected during commercial exploration for oil. The open ocean circulation in the area has also been well studied and modelled. In addition there are a considerable number of GPS referenced tide gauges providing accurate mean sea level at the UK coast. Finally it is known that UK geoid heights determined by levelling deviate too much from these tide gauge mean sea levels, a problem that is now overdue for resolution. Figure 1 shows the whole area where our geoid computation was performed, with region 1 containing the highest quality data. The location of GPS referenced tide gauges used for the validation studies is also shown.

Section 2 briefly describes the local land and marine gravity data available to us. Section 3 describes the method of geoid computation, which attempts to take account of longer and shorter wavelengths in the gravity data through a remove-restore procedure. Full details of this may be found in Hipkin *et al.* (2003). Section 4 describes the validation of the new geoid against tide-gauge and altimetric mean sea levels and dynamical model estimates of sea surface topography.



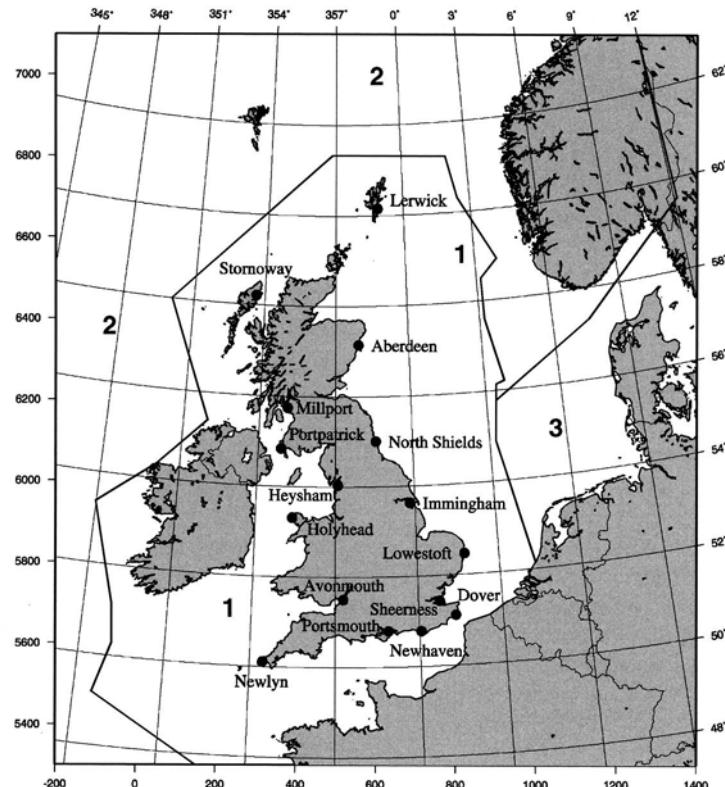


Figure 1. The area of geoid computation, also showing British tide gauge sites. The frame shows Universal Transverse Mercator (UTM), zone 30 co-ordinates in km (at left and bottom) and degrees latitude (right) and longitude (top). The three regions of marine gravity data (see section 2) are outlined.

Section 5 concludes with a discussion of prospects for applications with satellite measured gravity data.

2. Local Gravity Data Sources

The geoid calculation is based on a regular 1 km grid of gravity data in UTM zone 30 coordinates for the region shown in Figure 1. Three regions of marine data are marked in the figure indicating different sources and qualities of data. The best data were available in region 1. This was already gridded at 1 km by the British Geological Survey (BGS) as a consortium product with corresponding gravity agencies from adjacent countries. These best data came from an analysis in which ship track gravity had been smoothed along track before using crossover adjustments to obtain the final free air anomalies. This technique should be more widely used to obtain larger areas of consistently accurate ship gravity data. Bureau

Gravimetrique International (BGI, 1996) point data were used directly in region 3. Using raw data with no cross-over error adjustment will generate local irregularities in the geoid but they do not accumulate into longer wavelength trends because region 3 is small and is bounded by well-constrained data.

In region 2, KMS98 gravity, based directly on altimetric data (Andersen and Knudsen, 1998) were used. The absence of good quality *in situ* gravity in region 2 means we would not expect to retrieve good sea surface topography here. However the use of KMS98 data was found to give a smoother overall solution, with least impact across the boundary with the neighbouring region 1. In future it would clearly be desirable to develop integrated analyses, which could utilise different data types in different regions, directly as available.

Land gravity data were based on BGS data over the UK and BGI data over most of Europe. Data from Norway was not available so EGM96 free air anomalies (Lemoine *et al.*, 1997) were used here. Such a patched mosaic of *in situ* data is inevitable when calculating a local geoid, and from this perspective it should be possible to design suitable ways of including satellite measured gravity into an inversion procedure. More details of the treatment of the different data sets used in this analysis are given in Hipkin *et al.* (2003).

3. Computing the EDIN2000 Geoid

The objective was to derive a local marine geoid using methods capable of a 1cm accuracy over as large a part of the region as possible. This is the accuracy needed for practical tide-gauge and sea surface topography comparisons. The algorithmic methods used for the geoid inversion need to be capable of this level of accuracy although the quality of the input data may still determine the accuracy of the resulting geoid.

The algorithm used was a remove-restore procedure (as described in this section) exploiting the linear relationship between gravity and the geopotential, where components of the gravity are inverted to a geoid using different methods. The EGM96 gravity model, Lemoine *et al.* (1997) was used as a basic state, which provides geopotential coefficients in a spherical coordinate framework. Hipkin (2003) has shown how to calculate the self-consistent EGM96 gravity anomalies on an *ellipsoidal* Earth and these were initially subtracted from the measured grid of free air anomalies over our local region. The great advantage of including this very large scale ellipsoidal effect from the beginning is that the residuals of this process can then be treated with spherical approximations while retaining the required accuracy. This greatly simplifies the gravity to geoid inversions. To test how well this procedure removed the largest scale gravity variations within the region, the residual mean gravity anomaly over the whole region was calculated at 0.057mGals. This indicates that the ellipsoidal EGM96 gravity anomalies agreed with the measured ones on the largest scales to very great accuracy. In order not to

misrepresent this small mean residual it was removed from all the gridded gravity data in the region.

The second stage in the analysis was to convert the larger scale gravity residuals remaining in the data into a geoid component. This was done with spherical coordinates taking advantage of the fact that the ellipsoidal correction had already been removed. Spherical harmonic components between 20 and 200 were used to obtain the optimal projection of the gravity residuals, with the corresponding geoid components immediately available as spherical harmonic coefficients. After these gravity anomalies had been removed from the local data the remaining small scale residuals were very small. These were integrated in a planar geometry over the local region using a Fast Fourier Transform technique. The main purpose of the previous two stages was to achieve a final set of residuals that are sufficiently small and free of longer wavelengths, that they can be locally integrated using a greatly simplified Stokes kernel. Provided this is achieved the remove-restore process will ensure that the final geoid, consisting of the sum of the geoid components from the three stages, will be very accurate. Figures 2 and 3 show the large and small scale corrections to the EGM96 geoid which were identified in order to construct the final solution.

4. EDIN2000 Geoid Validation

4.1. TIDE GAUGES

First we consider tide gauge data. Suppose that the height of a benchmark is determined by levelling (an orthometric height, H) and, at the same benchmark, GPS measures the height above the ellipsoid, h . The difference is equal to the height of the geoid above the ellipsoid, N , plus the offset, Δ , of the levelling datum from the geoid (see Figure 4),

$$\Delta = h_s - H_s - N_s. \quad (1)$$

Equation 1 could include an adjustment for differential land uplift, from when the levelling was completed in the late 1950's and today, although the gravity data for our geoid also spans a considerable time period. The first test of a geoid is to compare N with GPS and levelling to see if Δ remains constant.

Inland in Britain, this method of testing geoid models is not generally available because commercial policy inhibits access to levelled heights. However at the tide gauges in Figure 1 we have both levelled and GPS ellipsoidal heights of the mean sea surface. With levelled heights for mean sea level, not only can we compare our geoid with levelling directly — but we can also study the deviations of the mean sea level from the geoid. Our knowledge of dynamics tells us that the coastal mean sea level should not deviate much from the geoid so this provides an additional constraint.

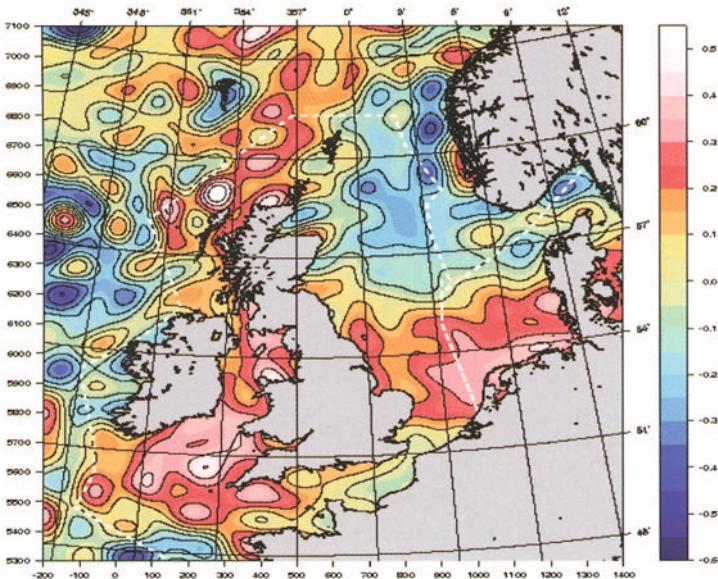


Figure 2. The large-scale additions to the EGM96 geoid. The RMS variability of the large-scale component is 0.33 m. The white dashed lines indicate the separation between the regions of different data in the EDIN 2000 calculation. Units are in m.

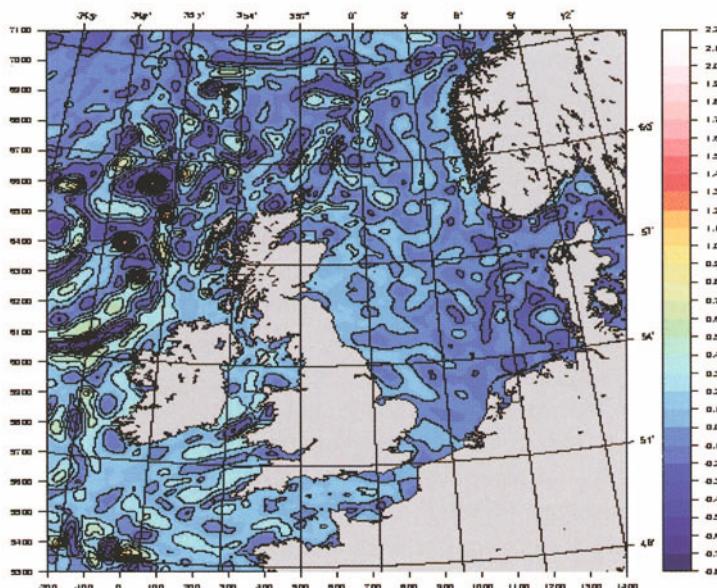


Figure 3. Shown are the small-scale additions to the EGM96 geoid. Together with the large-scale additions shown in Figure 2, the EDIN2000 geoid is made up. The RMS variability of the small-scale component is 0.14 m. Units are in m.

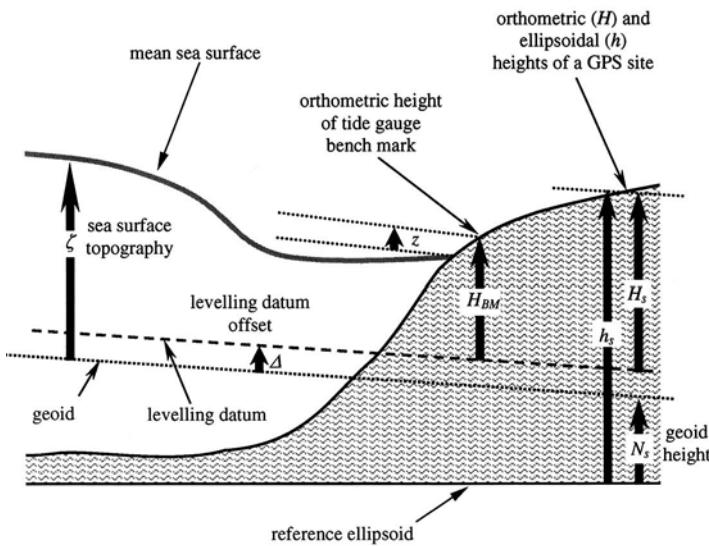


Figure 4. The geometry of height measurements near a tide gauge. Subscripts BM and s indicate quantities at the tide gauge benchmark and GPS station.

Figure 5a shows the levelled height of the mean sea surface at British tide gauges in Figure 1, above the referencing datum at Newlyn in Cornwall, that is $H_{BM} - z$ in Figure 4. To the north, the sea surface appears to be about 0.25 m higher than to the south, and there is the appearance of a step at about 53°N. These data are therefore similar to the results of Thompson (1980) who concluded that the sea level should not slope in this way relative to the geoid and that therefore levelling errors were present. Figure 5b shows ζ , the height of the mean sea surface, relative to geoid models, EDIN2000 and EGG97 (Denker and Torge, 1998). Each is reduced to zero mean but, for comparison, 0.5 m has been added to the EGG97 results.

Both geoids find the mean sea surface topography (MSST) to vary very little, much more in accord with oceanographic expectations than Figure 5a. The variability with EGG97 is 100 mm but with EDIN2000 it is considerably less (30 mm) if the local outliers Newlyn and Holyhead, (identified by open squares in Figure 5b and crosses in Figure 5c) are excluded.

Figure 5c shows the variability of the height of the levelling datum above the geoid, Δ in Figure 4. This compares the geoid and levelling directly, without reference to the mean sea surface, even though our orthographic heights happen to be restricted to tide gauges. Plotting Δ against latitude directly produces a step at 53°N as in Figure 5a which now must be definitely identified with an error in either the geoid or the levelling. Fitting the best constant values on either side quantified the step at 239.2 ± 53.6 mm. Figure 5c shows Δ or $\Delta - 239.2$ plotted against latitude with $\Delta - 239.2$ values indicated by open symbols. The standard deviations of Δ were 124.5 mm (EDIN2000) and 99.2 mm (EGG97) before correction and 43.7 mm (EDIN2000) and 61.0 mm (EGG97) afterwards. The fact that the vari-

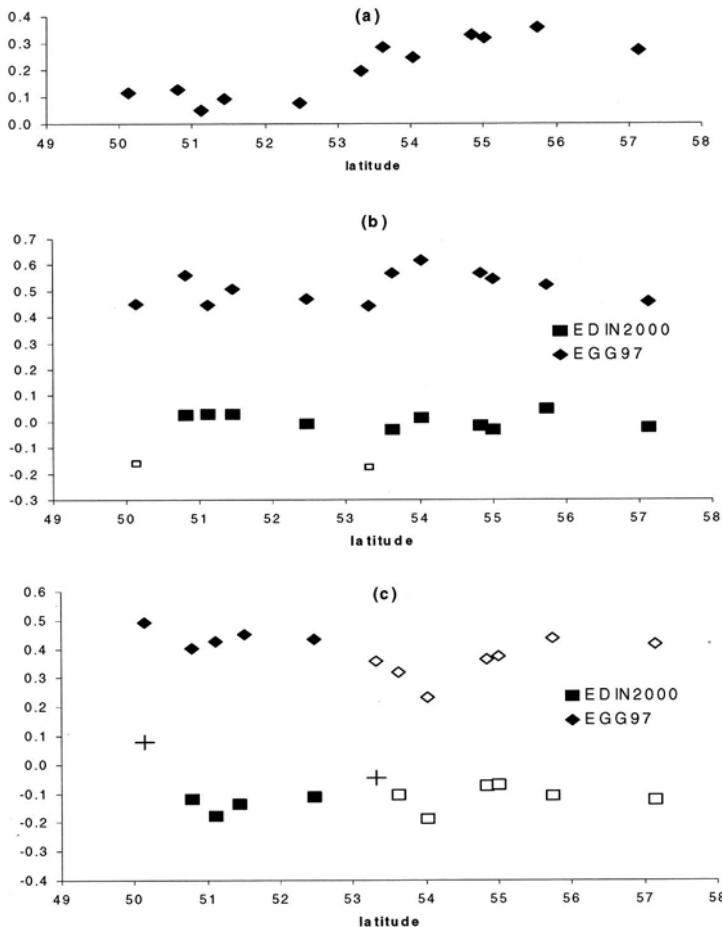


Figure 5. Height comparisons in metres at tide gauges. (a) Levelled height of the mean sea surface above Ordnance Datum Newlyn (ODN). (b) Height of the mean sea surface above the geoid of EDIN2000 and EGG97. Note the atypically large deviations at Newlyn and Holyhead for EDIN2000. These sites are shown in part (b) of the figure by open squares and by crosses in part (c). (c) Height of ODN above the geoid, after subtracting 239.2 mm from all levelled height north of 53°N. Corrected values are shown with open squares.

ability is larger in Δ than in ζ probably implies that the single step levelling error at 53°N is too simple.

4.2. SATELLITE ALTIMETRY

A model of the mean height of the sea surface has been prepared by CLS at Toulouse, using the altimeter satellites TOPEX/Poseidon, ERS-1 and Geosat (Hernandez and Shaeffer, 2000). The model (CLS01) represents an average shape of the ocean surface over the period 1993–99. Heights were provided on a regular grid of latitude and longitude with a spacing of $1/30^\circ$. These have been interpolated to the

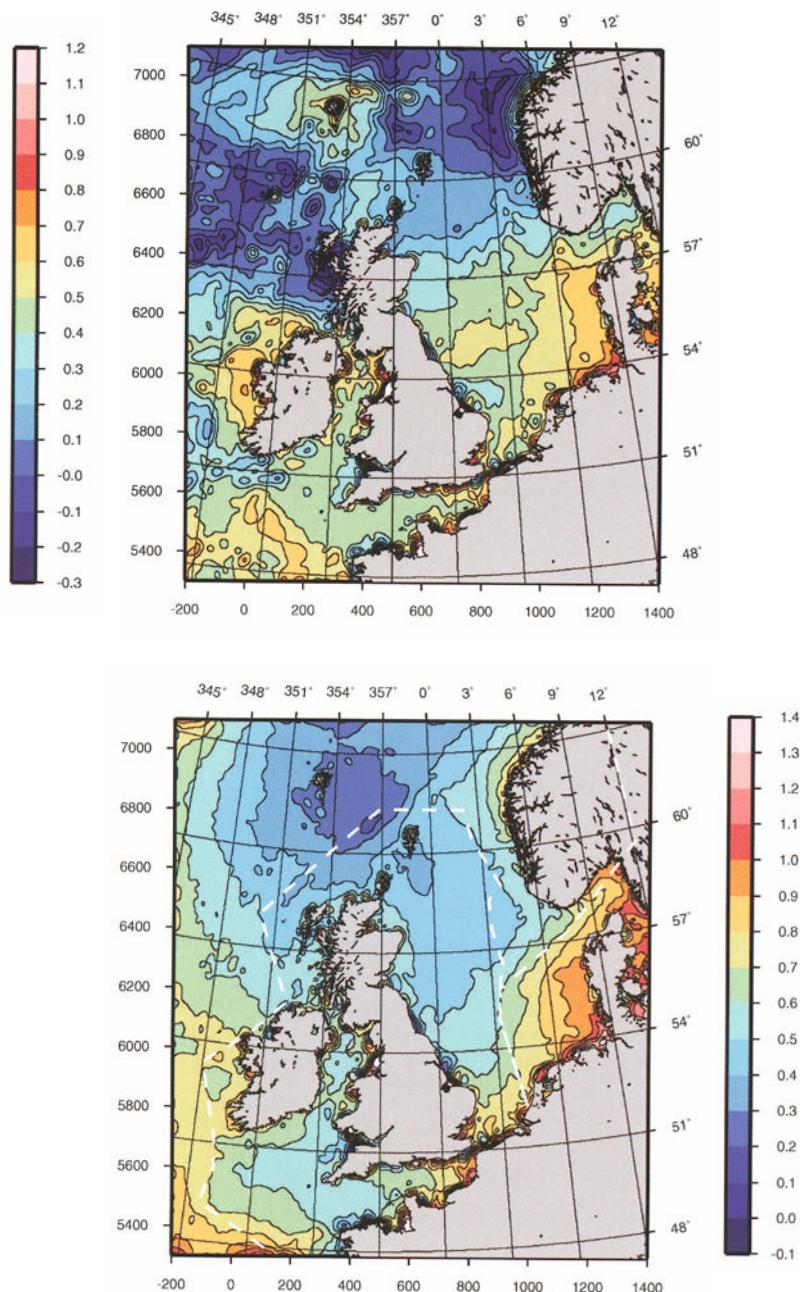


Figure 6. This Figure shows the mean sea surface topography calculated by CLS from altimeter sea level data by subtracting the EGG97 (top) or the EDIN2000 (bottom) geoids. The white dashed lines indicate the separation between the regions of different data in the EDIN2000 calculation. Any constant offset between fields may be ignored. Units are in m.

1 km UTM grid and both the EDIN2000 and EGG97 geoid models were subtracted to get two MSST estimates over the whole area. Figure 6 shows the estimates of MSST we obtain.

Clearly the mean dynamic topography below from the new geoid is considerably more realistic. The large N-S slopes in the northern North Sea and north of Ireland are certainly erroneous in the EGG97 Results. A number of features are worth noting:

1. The EDIN2000 MSST is generally very smooth where the gravity data are good over region 1 and corresponds to mean geostrophic velocities less than 100 mm s^{-1} .
2. Within region 1, there is a step in MSST across the shelf-edge, downwards away from north west Scotland and the Shetland Islands. It has an amplitude locally reaching 350 mm and is consistent in location with a north-eastward flowing shelf-edge current with a mean geostrophic velocity reaching 250 mm s^{-1} . There is evidence that some of this current diverts eastwards through the Pentland Firth into the North Sea.
3. We find only a very small change in sea surface topography when comparing the coast of northern England with that around Bergen in Norway, although there is a broad depression in the central North Sea of $100\text{--}150 \text{ mm}$.
4. The water piled up against the Danish and German coasts in region 3 agrees qualitatively with the known response of the North Sea to wind forcing, and the abrupt rise in MSST into the Kattegat is documented (Ekman and Makinen, 1996).
5. Beyond the pronounced step, MSST rises gently to the west of the shelf edge off the Northern Isles suggesting some weak currents towards the south. The large scale pattern in region 2 joins very consistently with region 1.

To put these MSST variations into context we now look at numerical ocean model results.

4.3. OCEAN MODELS OF MEAN SEA SURFACE TOPOGRAPHY

Figure 7 shows the MSST from two different ocean numerical models. The top model (Holt *et al.*, 2001) was developed by the Proudman Oceanographic Laboratory and was forced by UK Meteorological Office hourly wind stress and atmospheric pressure for one full year corresponding to 1989. This model is a limited area model and the entire domain is shown with tidal sea level variability used as forcing around the edge of the domain. The model below is the OCCAM global model (Webb *et al.*, 1998) on a $\frac{1}{4}$ degree grid of latitude and longitude. The model is forced by three-day averaged winds provided by the European Centre for Medium-range Weather Forecasting (ECMWF) for the period 1993–1995 inclusive. The model is also constrained by assimilated vertical profiles of water temperature measured with expendable bathy-thermographs, wherever these were available, and it should therefore have more accurate MSST slopes than an unconstrained model simulation (Fox and Haines, 2003).

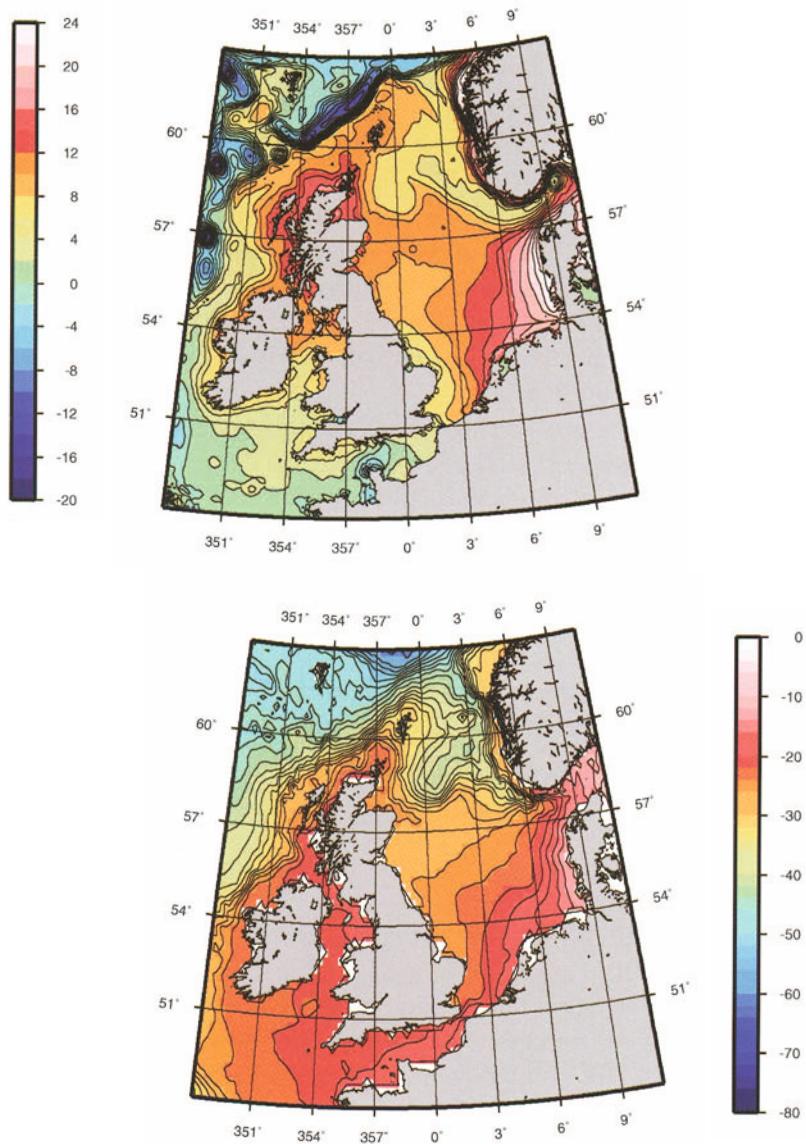


Figure 7. This Figure shows model mean sea surface topographies. The Proudman Shelf Seas model (top) for 1989, (in cm). The OCCAM model (bottom) for 1993–95 (in cm). Contour intervals 2 cm.

Both models show considerable consistency in the MSST over the continental shelf and both are quite consistent with the MSST derived directly from the altimetry and the EDIN2000 geoid data. Both models indicate a strong north-eastward current running along the shelf break where the continental shelf and deeper ocean meet. This current carries North Atlantic surface waters into the Norwegian Sea. The slope generates a step of about 20 cm downwards in a north-westerly direction in the region of the Shetlands, in good agreement with the EDIN2000 MSST in Figure 6. Both models indicate a cyclonic circulation in the northern North Sea although the strengths vary. Many of the other features of the altimeter-geoid MSST are also reproduced. The Proudman model is not realistic in the northwest corner away from the continental shelf because there are problems at the boundaries. Therefore the results from the large scale OCCAM model were included to show that the shelf edge current is a robust mean flow feature and not an artefact of the boundary forcing.

5. Conclusions

The main results of this paper are in the new method of calculating a local marine geoid of sufficient accuracy to resolve MSST. This result is robust to comparisons against tide gauge data and the best altimeter determined mean sea level. Two ocean models are used to calculate mean sea surface topography and the results are found to be consistent with the direct geoid based calculations.

The paper has shown that it is possible to calculate very accurate local geoids over quite large regions. We now need to see whether this accuracy can be maintained if the gravity data adjustment procedures used in region 1 are extended to the more sparse deep water gravity data away from the shelf edge. The region between the Shetlands and Iceland should provide sufficient ship-borne gravimetry to test these procedures and a new EU project GOCINA will try to produce an improved geoid across this whole region, which would then be of great importance for oceanographic studies.

Finally there will be a resurgence of interest in the calculation of improved regional geoids as a result of the new satellite gravity missions GRACE and GOCE. Geoids of the accuracy attained here will be very important for validating GOCE gravity field data and new techniques for melding surface, airborne and satellite determinations of the gravity field will be required. The methods developed here mark a significant step towards this goal.

References

- Andersen, O. B. and Knudsen, P.: 1998, 'Global marine gravity field from the ERS-1 and Geosat geodetic mission altimetry', *J. Geoph. Res.* **103C**, 8129.

- BGI (Bureau Gravimetrique International): 1996, CD-ROMs of BGI global gravity data, version 1.0, Bureau Gravimetrique International, Association Internationale de Geodesie, CNES, Toulouse.
- British Geological Survey: 1997, Colour shaded Relief Gravity Map of Britain, Ireland and adjacent areas: 1:1500000 scale, I.F. Smith & J.W.F. Edwards (compilers), Keyworth, Nottingham.
- Denker, H. and Torge, W.: 1998, 'The European gravimetric quasigeoid EGG97 — An IAG supported continental exercise', in Forsberg, Feissl, and Dietrich (eds.) *Geodesy on the Move: Gravity, Geoids, Geodynamics and Antarctica*, Springer, Berlin, 249–254.
- Ekman, M. and Makinen, J.: 1996, 'Mean sea surface topography in the Baltic and its transition into the North Sea: a geodetic solution and comparisons with oceanographic models', *J. Geoph. Res.* **101C**, 11993–11999.
- Fox, A.D. and Haines, K.: 2003, 'Interpretation of water transformations diagnosed from data assimilation', *J. Phys. Oceanogr.* **33**, 485–498.
- Hernandez, F. and Shaeffer, P.L.: 2000, *Altimetric Mean Sea Surfaces and Gravity anomaly maps intercomparisons*, AVI-NT-011-5242-CLS, CLS Ramonville St Agnes., 48 pp.
- Hipkin, R. G.: 2003, 'Ellipsoidal geoid computation', *J. Geod.*, in press.
- Hipkin, R. G., Haines, K., Beggan, C., Bingley, R., Hernandez, F., Holt, J., and Baker, T.: 2003, 'The Geoid EDIN2000 and Sea surface topography around the British Isles', *Geophys. J. Int.*, in press.
- Holt, J. T., James, I. D., and Jones, J. F.: 2001, 'An s-coordinate density evolving model of the northwest European continental shelf. Part 2: Seasonal currents and tides', *J. Geoph. Res.* **106C**, 14035–14053.
- Lemoine, F. G., Smith, D. E., Smith, R., Kunz, L., Pavlis, N. K., Klosko, S. M., Chinn, D. S., Torrence, M. H., Williamson, R. G., Cox, C. M., Rachlin, K. E., Wang, Y. M., Pavlis, E. C., Kenyon, S. C., Salman, R., Trimmer, R., Rapp, R. H., Nerem, R. S.: 1997, 'The development of the NASA GSFC and NIMA joint geopotential model EGM96', in *Gravity, Geoid and Marine Geodesy*, Springer, Berlin, pp. 461–469.
- Thompson, K. R.: 1980, 'An analysis of British monthly mean sea level', *Geoph. J. R. Astron. Soc.* **63**, 57–73.
- Webb, D. J., Coward, A. C., de Cuevas, B. A., and Gwilliam, C. S.: 1998, 'A multiprocessor ocean general circulation model using message passing', *J. Atmos. Ocean. Tech.* **14**, 175–183.

FEASIBILITY, AND CONTRIBUTION TO OCEAN CIRCULATION STUDIES, OF OCEAN BOTTOM PRESSURE DETERMINATION

CHRIS W. HUGHES (cwh@pol.ac.uk) and VLADIMIR STEPANOV (vst@pol.ac.uk)
Proudman Oceanographic Laboratory, Bidston Observatory, Prenton, Merseyside, CH43 7RA, U.K.

Received: 15 November 2002; Accepted in final form: 29 November 2002

Abstract. An assessment is presented of the probable magnitude of ocean signals causing aliasing in ocean bottom pressure measurements from the GRACE satellite mission. Even after modelling as much of the high frequency signal as possible, variability between 1 mbar (in quiet ocean regions) and 10 mbar (on some shelves) is likely to remain. Interpretation of the resulting retrievals will therefore rely on the facts that the satellite sampling will average the aliasing signal to some extent, and that the spatial patterns of aliased signal and true signal will be different. To this end, a theoretical argument is given, and supported by model diagnostics, suggesting that observable bottom pressure signals will be strongly constrained by the shape of the ocean floor. The modelled magnitudes offer the prospect of significant detectable signals and, while the model accuracy can be called into question, there are hints from Earth rotation and satellite orbit measurements that significant mass redistributions occur in the ocean. It seems certain that we will learn something new about the oceans from GRACE.

1. Introduction

The GRACE satellite mission, launched 17 March 2002, is expected to monitor temporal variations in the Earth's gravity field over a period of five years. These variations are the result of redistributions of mass in the atmosphere, ocean, hydro/cryosphere, together with long term adjustment processes in the lithosphere and mantle. Over the ocean, the mass variations are effectively a measure of ocean bottom pressure changes, averaged over some area. Assuming that such variations are slow, and that atmospheric pressure variations over land can be accurately modelled, it was shown by Wahr et al. (1998) that the expected accuracy is equivalent to 1 mm of water averaged over a Gaussian disc of radius 500 km. This is the only satellite technology currently capable of measuring anything other than a thin surface layer of the oceans.

In reality, ocean bottom pressure (and atmospheric pressure over land) varies significantly on time scales as short as hours, and not just the long time scales (periods longer than about 60 days) assumed by the idealized study. Since it takes typically 30 days to build up a global picture of the gravity field, and inversion of the gravity field to give mass distributions is a global procedure, any shorter period variations will produce aliased spatial patterns of apparent mass redistributions. These shorter period variations must be modelled, and their effect subtracted out of the observed quantities, in order to produce realistic maps of the long period vari-



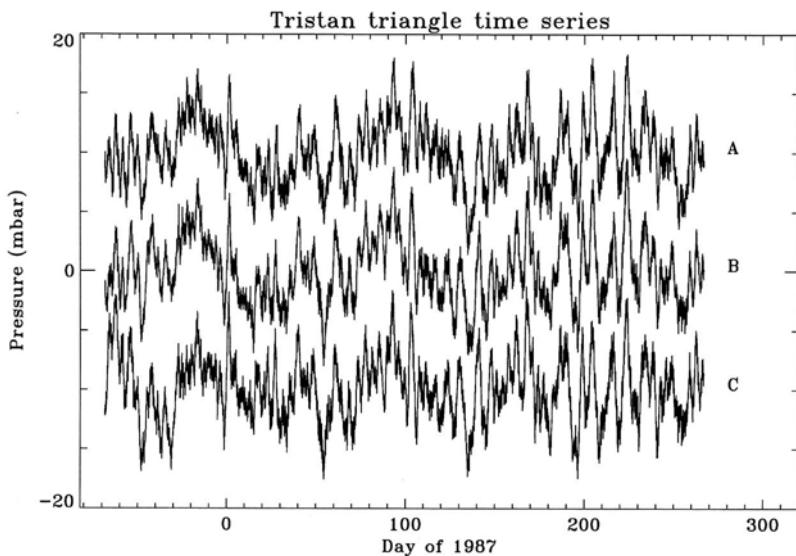


Figure 1. Time series of hourly ocean bottom pressures (tides removed), measured at three sites (A, B, and C) in an equilateral triangle about Tristan da Cunha island in the South Atlantic. Arbitrary offsets have been applied for visibility.

ations. This introduces a significant extra source of error in ocean bottom pressure estimates. The size of this error is hard to determine without detailed simulations, but it is clear that it will be given by the size of the unmodelled high frequency signal, multiplied by a mitigating factor less than 1 which represents the effects of averaging and filtering to take account of knowledge of the structure in time and space of the noise and expected signal.

In the next section, an assessment of the probable size of this noise is given. In section 3, a theoretical argument is given for the spatial structure of the observable long period changes in ocean bottom pressure, and the prediction is compared with output from a global, eddy-permitting ocean model.

2. Sources of Aliasing

Figure 1 shows hourly values (after subtraction of tides) of ocean bottom pressure from three bottom pressure recorders (BPRs) sited in triangle around the island of Tristan da Cunha in the S. Atlantic, in 1986/87. The three BPRs were separated by about 333 km (putting them at a distance of about 192 km from the triangle centre).

A large component of the bottom pressure signal is clearly coherent over length scales of a few hundred kilometres. A Wiener filter was used to extract this coherent part, by assuming that the differences between the time series represent the amplitude of the noise which causes a difference between a simple average of the

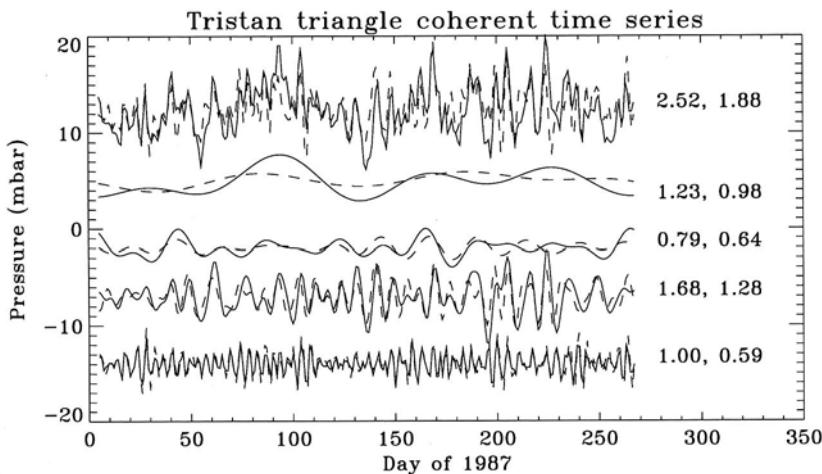


Figure 2. The coherent part of the Tristan bottom pressure timeseries shown in Fig. 1 (top, solid curve), and various band-pass filtered versions of that timeseries (periods > 60 days, 20–60 days, 6–20 days and 2–6 days, from top to bottom), shown with arbitrary offsets. Modelled versions of the same time series are superimposed (dashed lines). The numbers to the right of each curve are standard deviations of the measured time series (left), and of the residual after subtraction of model from measurements (right), measured in mbar.

three time series and a true area-averaged signal. The filter has amplitudes close to 1.0 at all frequencies and produces a time series of estimated area-averaged bottom pressure shown (with arbitrary offset) as the top, solid curve in Figure 2. Various band-pass filtered versions of this timeseries are also shown (other solid curves), together with their standard deviations. The part of the signal which is not coherent between pressure gauges represents a short length-scale signal which is undetectable by GRACE, and is not considered further. There is significant variability (around 1 mbar root-mean-square) at all frequency bands corresponding to periods of 2–6 days or longer, with 0.43 mbar at periods shorter than 2 days (including any residual tides after subtraction of a tidal fit). The bottom three curves thus represent the most troublesome signal for GRACE, and are dominated by the barotropic (depth independent) ocean response to wind stress and atmospheric pressure. The dashed line shows a preliminary model result from a global barotropic ocean model forced by analysed winds and pressures from the NCAR atmospheric model. With minimal tuning of the model, the root sum square from these three time series is reduced from 2.11 mbar to 1.55 mbar after subtraction of the modelled pressures (*i.e.*, squaring, summing and then taking the square root of the bottom three pairs of numbers in Fig. 2).

The Tristan site is a typical deep ocean site, with variability larger than most of the tropics, but smaller than certain “hot spots” in the Southern Ocean. More

detailed investigation is needed, but these initial results suggest that the high frequency (alias-producing) deep ocean signal can be reduced to below 2 mbar by applying model results.

More troublesome are the shallow shelf regions, particularly the large Siberian shelf. Signals over shelves are much larger and more sensitive to friction parameterisation. Experience of UK storm surge modelling shows that approximately 5 mbar accuracy can be achieved with careful attention to local conditions, but a global model is unlikely to achieve such accuracy. As with tides (Ray, this issue; Schrama, this issue), the unmodelled, high frequency signal over the shelf regions is likely to be a highly significant source of aliasing.

3. Expected Patterns

The large magnitude of these aliasing signals gives a rather pessimistic impression. However, it should be remembered that the pressure signal size is an upper limit on the final aliased signal: the complicated sampling by the satellites will produce varying degrees of averaging of the pressure, producing characteristic aliasing patterns. The trick will be to extract the unaliased signal from the combination of aliased and unaliased measurements by distinguishing their different spatial patterns. In doing this, some simple theory of ocean dynamics can be helpful. It is quite easy to show that, for slowly-changing ocean flows, the time dependent gradient of ocean bottom pressure must be closely aligned with the gradient of bottom topography.

Consider a current transporting T kgs $^{-1}$ northwards. The current is assumed to be distributed over some depth range h , and will be close to geostrophic balance if the oscillation period is much longer than 1 day, so

$$\rho f \hat{\mathbf{k}} \times \mathbf{u} = -\nabla_z p, \quad (1)$$

where ρ is water density, \mathbf{u} is the horizontal velocity vector, p is pressure (after subtraction of a horizontally-averaged pressure at each depth), ∇_z is horizontal gradient operator along a level surface, $\hat{\mathbf{k}}$ is the vertical (upwards) unit vector, and $f = 2\Omega \sin \theta$ where Ω is the Earth's angular rotation rate, and θ is latitude. This relationship relates the strength of the current to the pressure difference Δp across it. Dividing by f , integrating at constant z (height) across the current (width W) and multiplying by the vertical extent h of the current, the left hand side integrates to the mass transport T , giving this in terms of Δp :

$$T = \frac{h \Delta p}{f}, \quad \text{or} \quad \frac{\Delta p}{T} = \frac{f}{h} \quad (2)$$

thus we have a scaling for pressure differences in the ocean. Were there no other dynamical constraints, this could be expected to be a valid scaling for variations in ocean bottom pressure (after subtracting the time mean hydrostatic component).

However, the depth-integrated vorticity balance provides a very strong dynamical constraint. Integrating (1) over the depth of the ocean from $z = -H$ (ocean floor) to $z = \eta$ (the surface), and then taking the (vertical component of the) curl (and assuming the horizontal divergence of depth-integrated mass flux is zero, meaning there are no mass sources) gives

$$\beta V = \nabla_h p_b \times \nabla_h H, \quad (3)$$

where $\beta = \partial f / \partial y$, with y being northward distance, p_b is pressure at the ocean floor, V is the vertical integral of ρv , where v is the northward component of \mathbf{u} , and ∇_h is the horizontal gradient of a 2-dimensional quantity. The more general version of (3) is given by Hughes and de Cuevas (2001), where wind forcing and other terms are also considered and it is shown that (3) is a good approximation for the large-scale ocean flow, including western boundary currents.

The significance of (3) is that it places a strong constraint on the variation of pressure along a depth contour. Writing s as the coordinate measured along a depth contour, and n as the perpendicular coordinate, (3) becomes

$$\beta V = \frac{\partial p_b}{\partial s} \frac{\partial H}{\partial n}, \quad (4)$$

and a horizontal integral across the current, in the direction of n , gives

$$\beta T = \frac{\partial p_b}{\partial s} \Delta H, \quad (5)$$

where ΔH is the difference in ocean depth across the current (V is depth-integrated mass transport, and T is depth- and horizontally-integrated mass transport).

Equation (2) tells us the pressure difference Δp across a current carrying transport T , and now (5) tells us the pressure gradient along a depth contour necessary for vorticity balance of transport T . There will therefore be some distance L along a depth contour over which the pressure gradient integrates to Δp . From (5), this is given by

$$\beta T = \frac{\Delta p \Delta H}{L}, \quad \text{or} \quad \frac{\Delta p}{T} = \frac{L\beta}{\Delta H}. \quad (6)$$

Combining (2) and (6), and noting that $\beta = f/(R \tan \theta)$ where R is the Earth's radius, we obtain an equation for L :

$$L = \frac{\Delta H}{h} R \tan \theta. \quad (7)$$

This means that L is the same order as, or larger than, the distance from the equator ($R\theta$), unless $\Delta H \ll h$, which would mean that the vertical extent of the current is much smaller than the range of ocean depths spanned by its lateral extent – a rather unusual circumstance.

We should therefore expect pressure gradients perpendicular to depth contours to be larger than those along the contours by a factor of order R/W , where R is the Earth's radius and W the current width. For deep or bottom-flowing currents, W is constrained by topography to be much less than R . Ocean bottom pressure variations should therefore be dominated by patterns that closely follow depth contours. In fact this is understating the case, since it is possible for much larger pressure gradients across depth contours to occur with no associated transport, if the water below a certain depth simply changes density (changed hydrostatic balance producing a change in bottom pressure but no geostrophic flow).

In physical terms the reason for this scaling is the constraint rotation places on vertical stretching of the water column. A pressure gradient along a depth contour implies a geostrophic flow up or down the slope, and therefore a vertical velocity (which must decrease to zero, or a small value, at the surface). To avoid this stretching, and thereby spinning up the water column, it is easiest for water to flow along depth contours.

This prediction is verified by diagnostics from a global, eddy-permitting ocean model OCCAM (Ocean Circulation and Climate Advanced Modelling project), described in Killworth (1995). The model was spun-up by relaxing the temperature and salinity fields to climatological values, and applying climatological wind stress, until the eddy field was fully developed. The temperature and salinity relaxation was then turned off, but surface wind stress continued to be monthly climatological values. The model (like the real ocean) was not in thermohaline equilibrium so, after the spin-up, slow changes in density and deep circulation developed. Although unrealistic in detail, such "model drift" can be used as a guide to the way the ocean would respond to interannual differences in deep water formation (the model drift is thought to be mostly the result of the lack of a sea ice model, and therefore lack of deep water formation).

An average of bottom pressure was calculated for each of two years (years 1 and 4 after spin-up of the model). The difference between these two pressure fields is shown in Fig. 3. Strong signals are seen in the Southern Ocean and on continental shelves (although the latter in particular may be a model artifact). In all regions, however, it is clear that the pattern of bottom pressure change is strongly constrained by topography (contoured), with the mid-ocean ridge system particularly important.

It could be argued that the signals are exaggerated in this model, which has not had time to approach thermodynamic equilibrium. However, there are intriguing hints that significant mass redistribution does occur in the real ocean. Hide and Dickey (1991) noted a fluctuation in the Earth's rotation rate associated with the 1982–83 El Niño, even after accounting for the atmospheric contribution to the signal, hinting at a significant role for the ocean. More recently, Cox and Chao (2002) noted a dramatic change in the J_2 coefficient of the Earth's gravity field coincident with the end of the 1997–98 El Niño, implying a mass redistribution in the oceans, hydrosphere, or outer core.

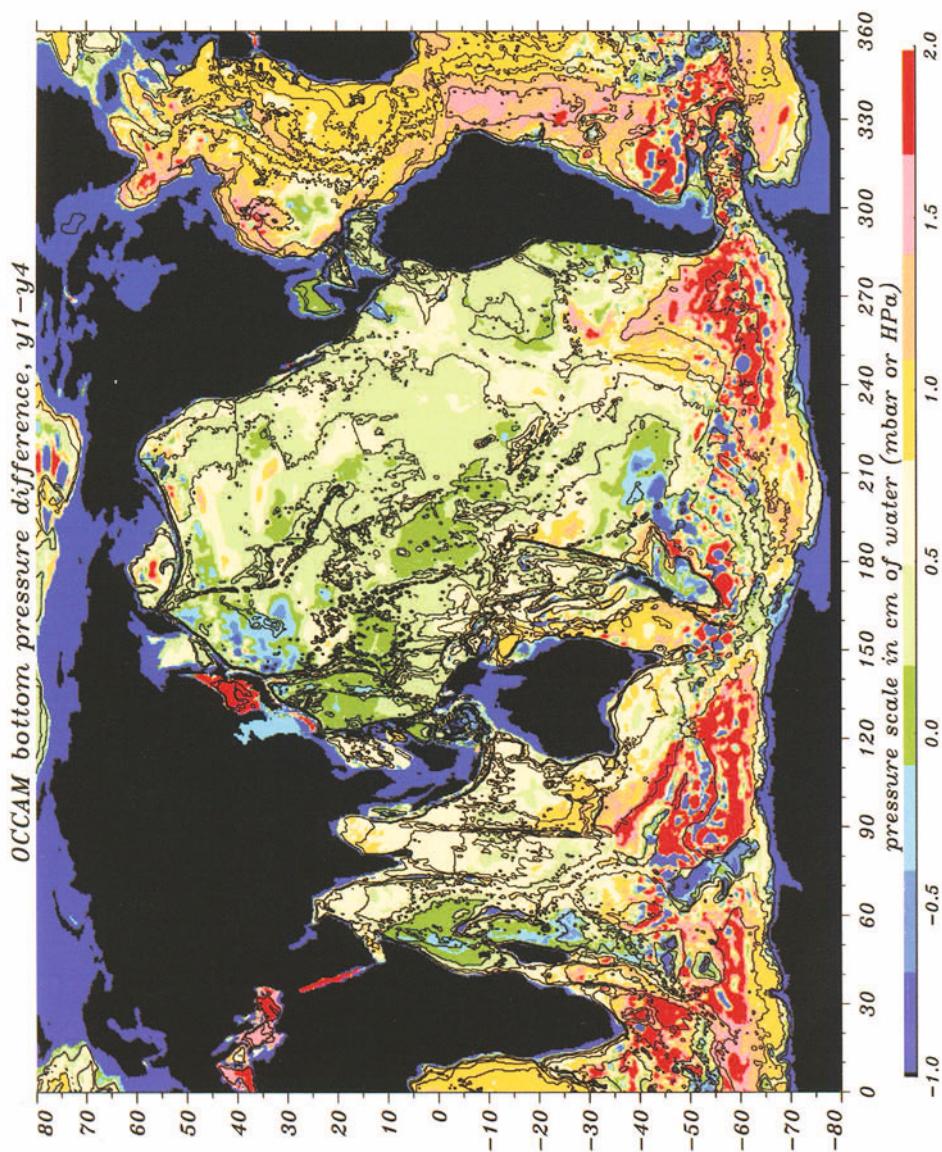


Figure 3. The difference between two annual average bottom pressure fields from the OCCAM model, with depth contours (2, 3, 4, and 5 km) contoured.

4. Conclusion

Even after the best efforts at modelling the fast ocean response to wind stress and atmospheric pressure forcing, it is likely that errors will remain of the order of 2 mbar in the open ocean, and larger over the shelves. These will produce somewhat smaller aliased patterns of apparent mass distribution in GRACE retrievals. Distinguishing these aliased patterns from ocean dynamics will rely on being able to recognise the different characteristic patterns of the two kinds of signal. To this end, it is of interest to note that theoretical arguments and model diagnostics both support the idea that the long period changes in ocean bottom pressure will be strongly constrained by the shape of the bottom topography.

Model results predict long period signals of up to 5 mbar in the Southern Ocean, on basin scales easily resolved by GRACE, and 1 to 2 mbar elsewhere. While it may be reasonable to call into question the validity of the model on these time scales, there are strong hints from Earth rotation and satellite tracking data that significant mass redistributions in the ocean accompany El Niño. It is hard to predict what it will be, but it seems certain that we will learn something important about the deep ocean circulation when GRACE data are processed.

References

- Cox, C. M., and B. F. Chao, 2002: Detection of a large-scale mass redistribution in the terrestrial system since 1998. *Science*, **297**, 831–883
- Hide, R., and J. O. Dickey, 1991: Earth's variable rotation. *Science*, **253**, 629–637
- Hughes, C. W., and B. A. de Cuevas, 2001: Why western boundary currents in realistic oceans are inviscid: A link between form stress and bottom pressure torques. *J. Phys. Oceanogr.*, **31**, 2871–2885
- Killworth, P. D., 1995: OCCAM: the Ocean Circulation and Climate Advanced Modelling Project. *International WOCE Newsletter*, **18**, 20–21 (unpublished manuscript).
- Wahr J., M. Molenaar, and F. Bryan, 1998: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.*, **103**(B12), 30205–30229
- Address for Offprints:* Chris W. Hughes, Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, U.K.

IMPACT OF GEOID IMPROVEMENT ON OCEAN MASS AND HEAT TRANSPORT ESTIMATES

P. LE GRAND (p.legrand@ifremer.fr)
IFREMER, BP 70, F-29280 Plouzane, France

Received: 17 May 2002; Accepted in final form: 15 November 2002

Abstract. One long-standing difficulty in estimating the large-scale ocean circulation is the inability to observe absolute current velocities. Both conventional hydrographic measurements and altimetric measurements provide observations of currents relative to an unknown velocity at a reference depth in the case of hydrographic data, and relative to mean currents calculated over some averaging period in the case of altimetric data. Space gravity missions together with altimetric observations have the potential to overcome this difficulty by providing absolute estimates of the velocity of surface oceanic currents. The absolute surface velocity estimates will in turn provide the reference level velocities that are necessary to compute absolute velocities at any depth level from hydrographic data.

Several studies have been carried out to quantify the improvements expected from ongoing and future space gravity missions. The results of these studies in terms of volume flux estimates (transport of water masses) and heat flux estimates (transport of heat by the ocean) are reviewed in this paper. The studies are based on ocean inverse modeling techniques that derive impact estimates solely from the geoid error budgets of forthcoming space gravity missions. Despite some differences in the assumptions made, the inverse modeling calculations all point to significant improvements in estimates of oceanic fluxes. These improvements, measured in terms of reductions of uncertainties, are expected to be as large as a factor of 2.

New developments in autonomous ocean observing systems will complement the developments expected from space gravity missions. The synergies of in situ and satellite observing systems are considered in the conclusion of this paper.

1. Introduction

As it is more and more evident that human activities have a significant impact on the climate system (IPCC, 2001), it becomes highly desirable to improve our capability to observe, quantify, and predict climate change. The ocean is an important component of the climate system through its role in redistributing heat from equatorial regions to polar regions (Ganachaud and Wunsch, 2000) and through its role in regulating the carbon cycle (Sigman and Boyle, 2000). In turn, climate change can potentially affect the ocean circulation and chemistry through various feedback mechanisms (Sigman and Boyle, 2000). Observations of the ocean have until recently been scarce, not the least because of the difficulty of collecting data in this high-pressure, corrosive, opaque, and generally hostile environment. The past decade, however, has seen the advent of precise satellite observing systems like space-borne altimeters, which for the first time have allowed physical oceanogra-



phers to observe the ocean circulation in a global and quasi-synoptic way. Two major difficulties remain, however, in using altimetric data to quantify the ocean circulation and the processes that are associated to it. The first difficulty is that altimetric observations only sample the surface of the ocean. In situ measurements are therefore required to constrain the interior, the circulation being highly variable from the surface to the bottom of the ocean. In the oceanographic terminology, the ocean circulation is said to contain a strong baroclinic, or non-barotropic, component. The interpretation of hydrographic measurements in terms of baroclinic circulation are developed in section 2 of this paper.

The second difficulty in using altimetric data to quantify the ocean circulation is that these data do not constrain the ocean circulation directly. Instead, they constrain the sum of the geoid height and of the ocean dynamic topography, so it is difficult to disentangle the two signals. This problem is reviewed in section 3.

In the near future, dedicated space gravity missions like GOCE and GRACE will allow a precise determination of the geoid height and will thus allow the separation of the ocean circulation signal from the geoid signal in altimetric observations. Resulting improvements in circulation estimates are reviewed in section 4 of this paper, with a particular attention on volume and heat transports that are two quantities of great importance in the climate system. The last section of this paper concludes by investigating the complementarities between observations provided by the forthcoming space gravity missions and new automated in situ measurement techniques that are now becoming available on the global scale. The complementarities of the GOCE and the GRACE gravity missions are also considered in this section.

2. Hydrographic Measurements and Ocean Circulation

In situ measurements that constrain the baroclinic ocean circulation are available in the form of hydrographic observations. These observations consist of temperature and salinity measurements collected at a known pressure level using a Conductivity, Temperature, Depth (CTD) instrument lowered from an oceanographic vessel. The data are thus collected as vertical profiles at a station above a more or less fixed point. Using the equation of state of sea water (Fofonoff, 1985), temperature and salinity are then converted into density. Density is in turn related to the baroclinic ocean circulation through the thermal wind relations:

$$f \frac{\partial u}{\partial z} = + \frac{g}{\rho} \frac{\partial \rho}{\partial y} \quad ; \quad f \frac{\partial v}{\partial z} = - \frac{g}{\rho} \frac{\partial \rho}{\partial x} \quad (1)$$

where u and v are the horizontal components of the velocity field, f is the coriolis parameter, ρ is the density, g is the gravitational acceleration, x and y the horizontal coordinates, and z is the vertical coordinate. These equations are derived (see Pond and Pickard, 1983) from 1) the geostrophic balance, which assumes

that the dominant terms in the horizontal momentum equations are the Coriolis force and the pressure force 2) the hydrostatic balance, which assumes that vertical gradients of pressure are proportional to density. These approximations are very good ones in the ocean, and deviations from them are in most areas unmeasurably slight (Wunsch, 1996). The main exceptions, where the approximations completely fail, occur at the equator where the Coriolis parameter vanishes and near the surface of the ocean where a large fraction of the velocity field is directly caused by the transfer of momentum from the atmosphere to the upper ocean. The terminology “thermal wind” comes from meteorology because in the atmosphere density can be replaced by temperature using the law of ideal gases. Note that, because of the simplifications made, the thermal wind equations, as well as the geostrophic and hydrostatic equations, are not self-contained. Indeed, unlike the full-blown Navier-Stokes equations, which can in principle be solved if initial/boundary conditions are known, the thermal wind relations require observational information in order to determine the ocean circulation. For this reason, the thermal wind equations are said to be diagnostic as opposed to the Navier-Stokes equations, which are said to be prognostic. Integrating (1) with respect to z , one obtains:

$$u(h) = u_r + \int_{z_r}^h \frac{g}{f\rho} \frac{\partial \rho}{\partial y} dz \quad ; \quad v(h) = v_r + \int_{z_r}^h \frac{-g}{f\rho} \frac{\partial \rho}{\partial x} dz \quad (2)$$

where $u(h)$ and $v(h)$ are the horizontal velocities at any depth level h , and u_r and v_r are the velocities at the arbitrarily chosen depth z_r . One sees from (2) that horizontal velocities at depth level h are determined from the density of sea water, up to the arbitrary constants u_r and v_r . These constants are usually called reference level velocities, and z_r is called the reference level. In the early days of oceanography, it was common to make the assumption that, provided that the reference level was chosen deep enough or in between water masses that flow in opposite directions, the reference level velocities could be neglected. In the 1970s, this traditional assumption was shown to be far from valid, and more rigorous approaches had to be designed. This led to the introduction of inverse methods in oceanography (Wunsch, 1978) which use available information to estimate the reference level velocities instead of neglecting them. This available information is quite eclectic (Wunsch, 1984). It comes, for instance, from the constraint that mass should be conserved within closed volumes in the ocean. If one considers a trans-Atlantic hydrographic section at 24°N, for instance, the reference level velocities must be consistent with the constraint that the net flux of mass across this section is close to zero. Indeed, any large flux would have to be balanced by large sources or sinks of water that are not observed north of the section (in this example, it is assumed, for the sake of simplicity, that the net flux of water across the Bering Strait and the net air-sea flux of water over the North Atlantic are negligible). If the oceanic section is discretized onto a number of gridpoints, the constraint of zero

net flux can be formalized as a set of linear equations that relates discrete reference level velocities to the observations:

$$\mathbf{Ax} = \mathbf{b} \quad (3)$$

where \mathbf{A} is a known matrix, \mathbf{x} is a vector containing the unknown reference level velocities at each gridpoint, and \mathbf{b} is a known vector containing zeros or, more generally, quantities known from observations. Most often, the number of equations is smaller than the number of unknowns in (3), so the problem is underdetermined. Moreover, the equations contained in (3) are never fully exact because of discretization errors and, more importantly, because of observational uncertainties. For the mass conservation equation associated with the 24°N section, for instance, the net flux of water across the Bering Strait is known to be small but not exactly zero (Coachman and Aagaard, 1988), so one cannot strictly impose that the flux of water across 24°N is zero. Formally, one should therefore replace (3) by the following equation:

$$\mathbf{Ax} = \mathbf{b} \pm \mathbf{n} \quad (4)$$

where \mathbf{n} is a vector containing the unknown errors in the constraints. Thus, (4) is always underdetermined. It can still be solved, however, using some variant of least-squares techniques that, for instance, minimizes the departure from a known prior estimate. Obviously, the least-squares solution will be subject to uncertainties, but these uncertainties can also be computed as part of the least-squares procedure so that one can attribute a degree of confidence to the solution.

Once the reference level velocities have been estimated, the oceanic transports of volume, or of any other property, can be computed using (2). The ocean inverse estimation procedure can also be implemented over a gridded model. In this case, the mass conservation constraints, or any other type of constraints, are written locally for each box defined by the grid of the model. The available constraints are numerous, ranging from dynamical constraints such as mass and heat conservation, to observational constraints other than hydrographic observations (current meter moorings that measure velocities at a fixed point in the ocean, for instance). Satellite altimetry is one of the newest, and potentially one of the most powerful, observational constraints. How these remote sensing observations can be implemented in the eclectic inverse procedure described above is the focus of the following section.

3. Altimetric Constraints and the Ocean Circulation Estimation Problem

Altimetric observations constrain the sea surface height, which is the sum of 1) the ocean dynamic topography caused by surface oceanic currents and 2) the geoid height that corresponds to the elevation the sea surface would have if the ocean were at rest. Currently available estimates of the geoid height, which is by far

the dominant signal, are not precise (errors of several 10's of cm at the spatial scales of 1 degree that characterize intense oceanic currents such as the Gulf Stream (Lemoine *et al.*, 1998)). It is thus difficult to remove the geoid signal from the sea surface height observations in order to precisely estimate the residual dynamic topography, which is the quantity of interest to physical oceanographers. Note, however, that the time-varying component of the ocean circulation can be determined because temporal variations in the Earth gravity field are negligible compared to temporal variations in the ocean dynamic topography. Altimetric observations are easily implemented in the inverse estimation procedure. Indeed, these observations provide constraints on velocities at the surface of the ocean through the geostrophic relations:

$$fu(0) = -g \frac{\partial \eta}{\partial y} \quad ; \quad fv(0) = +g \frac{\partial \eta}{\partial x} \quad (5)$$

where η is the ocean dynamic topography and $u(0)$ and $v(0)$ are the geostrophic velocities at the surface of the ocean. Non-geostrophic Ekman velocities caused by the direct effect of wind forcing are also present at the surface of the ocean (Pond and Pickard, 1983). It is assumed in the following that the volume transport associated with these velocities can be independantly estimated so that it does not limit the precision of overall transport estimates. Although this assumption is not valid at present, there is good hope that accurate ocean surface wind fields will become available in the near future through the assimilation of satellite scatterometer observations in atmospheric general circulation models (see the Question and Answer section in Milliff *et al.* (2001) for a discussion of this point). Note that the improvements in transport estimates presented below would largely remain valid for the geostrophic component of the velocity field even in the absence of improvements in the Ekman component.

A discretized version of equation (5) provides observational constraints on surface velocities which can easily be added to the set of linear equations (4). Because they are mostly due to errors in the geoid height model that are spatially correlated, uncertainties in the altimetric constraint must be represented by a full covariance matrix. At present, this constraint moderately improves the inverse solution, mostly in the Southern Ocean (LeGrand *et al.*, 2002). It has not yet, however, reached its full potential. It will do so when precise geoid models provided by space gravity missions like GOCE and GRACE, and to some extent CHAMP, become available. The expected improvements are reviewed in the following section.

4. Impact of Geoid Improvement on Ocean Mass and Heat Transport Estimates

With improved estimates of the geoid, the conversion of altimetric data into absolute dynamic topography will become possible. Using this information, oceanographers will be able to precisely estimate the circulation at the surface of the ocean.

These improved surface estimates, combined with in situ hydrographic measurements, will in turn provide strong constraints on the ocean circulation at all depths. The resulting improvements can be quantified in terms of reduction of the uncertainties in the least-squares solution of (4) without knowing the actual geoid height provided by the gravity missions.

Following this approach, improvements in oceanic transport estimates have been quantified in three inverse modelling studies using the expected precision of future gravity missions. Two of them (Ganachaud *et al.*, 1997; Schröter *et al.*, 2002) use box inverse models in which oceanic transports are calculated across trans-oceanic sections, whereas the third one (LeGrand, 2001) uses a finite difference inverse model. Wunsch and Stammer (this volume) also show how improved geoid measurements can be used in an ocean data assimilation framework to improve both estimates of the ocean circulation and the geoid itself. Inverse modelling and data assimilation techniques are conceptually very similar, although a stronger emphasis is put on the dynamical constraints in the second approach. The most notable advantage of considering a more complex dynamical framework is the inclusion of time-dependent processes. For that reason, the data assimilation approach will allow the complete interpretation of the time-dependent observations of the gravity field provided by the GRACE mission. (see the article by Nerem in this volume for a discussion of the time-varying effects resolved by GRACE.) The approach described here has the advantage of being easy to implement, and to provide estimates of the uncertainty in the solution.

The Ganachaud *et al.* (1997) calculation studies the impact of three different hypothetical low-low satellite-to-satellite tracking missions, the specifications of the third mission including GPS tracking and resembling the specifications of the GRACE mission. In Ganachaud *et al.*'s (1997) calculation, only the reference velocities are treated as variables and therefore the velocity uncertainties due to the terms of equations (2) proportional to the ocean density field are not explicitly taken into account. In the calculation corresponding to a GRACE-like mission (specifications provided for a spherical harmonic expansion up to degree 70), the density uncertainties are set to zero in order to explore the best possible outcome. For that reason, the results of Ganachaud *et al.* (1997) shown in Table I can be considered as an optimistic estimate of the impact of GRACE based on the assumption that all errors in the in situ density field could be rendered negligible by the use of a network of simultaneous and precise in situ observations. These optimistic assumptions are not justified if the objective is to estimate transport uncertainties in layers of the ocean because transports calculated over a limited depth range are often dominated by the density term of equation (2) and thus the uncertainties in these transports are controlled by the uncertainties in the density field. The assumptions can be partly justified, however, when the objective is to estimate the uncertainties in transports integrated from the surface to the bottom of the ocean. Indeed, reference level velocities often dominate vertically integrated transports across trans-oceanic sections because, despite being small, they translate

TABLE I
Results of various impact studies.

Study	Ganachaud <i>et al.</i> (1997) (heat fluxes, Bettadpur 70 case)	LeGrand (2001) (Vol- ume fluxes in the upper km of the ocean)	Schröter <i>et al.</i> (2002) (heat flux in section across the Gulf Stream at 24°N)
Transport uncertainty reduction	1% to 60%	7% to 42% for GOCE and 2% to 20% for GRACE	37% for GOCE and 29% for GRACE

into large transports when integrated over the whole water column. Note that the uncertainties in volume transports across trans-oceanic sections closing an ocean basin are likely to be very small, even without the altimetric/geoid constraint on dynamic topography. Indeed, volume transports across these sections cannot significantly deviate from zero, as in the example of the 24°N section, because there is no large net flux of volume across the northern boundary of the basins closed by these sections. It is therefore not surprising that Ganachaud *et al.* (1997) find that mass* transport uncertainties are usually not reduced much for the sections that close an ocean basin. Accordingly, they find that the largest impact occurs in the open Southern Ocean basin. The same argument cannot be applied to the case of heat transports because these transports are largely caused by the different heat contents of surface waters flowing in one direction and deep waters flowing in the opposite direction (Figure 1). Thus, heat transports can be large across a section, even if the corresponding net volume transport is equal to zero. Ganachaud *et al.* (1997) find a reduction in heat flux uncertainties that range between negligible values in some areas of the North Atlantic and values of 60% in the South Pacific/Indian Oceans and in the Circumpolar Current (Table I).

A different impact calculation is presented in LeGrand (2001). In this study, the uncertainties in the density field are explicitly taken into account so that the impact of gravity missions on transport uncertainties can be estimated for various layers of the ocean. The impact of the GOCE and GRACE missions is estimated across various sections of the Atlantic Ocean, including sections across narrow oceanic currents like the Gulf Stream, the Azores Current, and the Benguela Current. The Atlantic Ocean contains most of the processes active in the ocean, including deep water formation, boundary currents, and wind driven gyres so the results of the study are fairly general despite the limited geographical extent of the model domain. Because temperature is not an explicit variable in the inverse model, only the

* Mass transports and volume transports describe the same processes. Mass transports are more relevant because mass is strictly conserved in the ocean but not volume. However, replacing the condition of non-divergence of mass fluxes by a condition of non-divergence of volume fluxes is a good approximation almost everywhere.

OVERTURNING CELL IN THE ATLANTIC

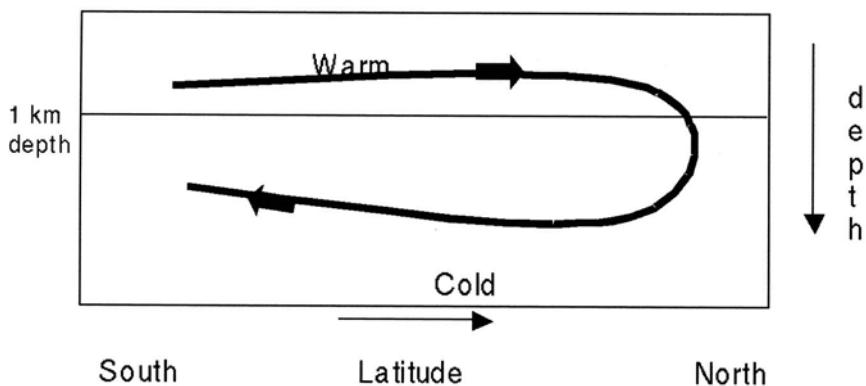


Figure 1. Schematic representation of the overturning cell in the Atlantic.

impact of gravity missions on volume fluxes is investigated. However, meridional heat fluxes in the Atlantic are conditioned to a large extent by the overturning cell (Hall and Bryden, 1982) by which the northward flow of warm waters in the upper layers of the ocean is compensated by a return flow of cold waters in the deep layers of the ocean (Figure 1). Thus, by looking at the impact of gravity missions on volume fluxes in the upper kilometer or so, one gets a good idea of how heat fluxes will be improved. The uncertainty reductions obtained by LeGrand (2001) are shown in Figure 2 for the upper kilometer of the ocean, and summarized in Table I. In the GOCE calculation, they range from small improvements across trans-oceanic sections to improvements larger than 40% in the Brazil Current and in the Circumpolar Current. In absolute terms, the uncertainty reductions reach several Sverdrups (1 Sverdrup = $10^6 \text{ m}^3/\text{s}$). In the GRACE calculation, the impact is, on average, slightly less than half the impact of GOCE because GRACE will not resolve the small spatial scales of the ocean circulation. Note that the inverse calculation is a climatological one so the temporal variations that GRACE intends to observe are not investigated. The main conclusion of the study is that, when the uncertainties in the density field are taken into account, the impact of gravity missions on oceanic transport estimates is smaller at depths. This result is not surprising since the combination of altimetry and gravity observations will provide information on the circulation at the surface of the ocean, and the transmission of this information to the deep layers of the ocean will be hampered by the presence of noise in the in situ observations of the density field. The main contribution of gravity missions will thus be an improvement in transport estimates in the upper branch of the meridional overturning cell shown schematically in Figure 1.

The box-inverse model impact study of Ganachaud *et al.* (1997) has been revisited by Schröter *et al.* (2002) in view of the characteristics recently specified for GOCE and GRACE. Moreover, Schröter *et al.* (2002) investigate the potential impact of gravity missions across a trans-oceanic section, the 24°N one mentioned

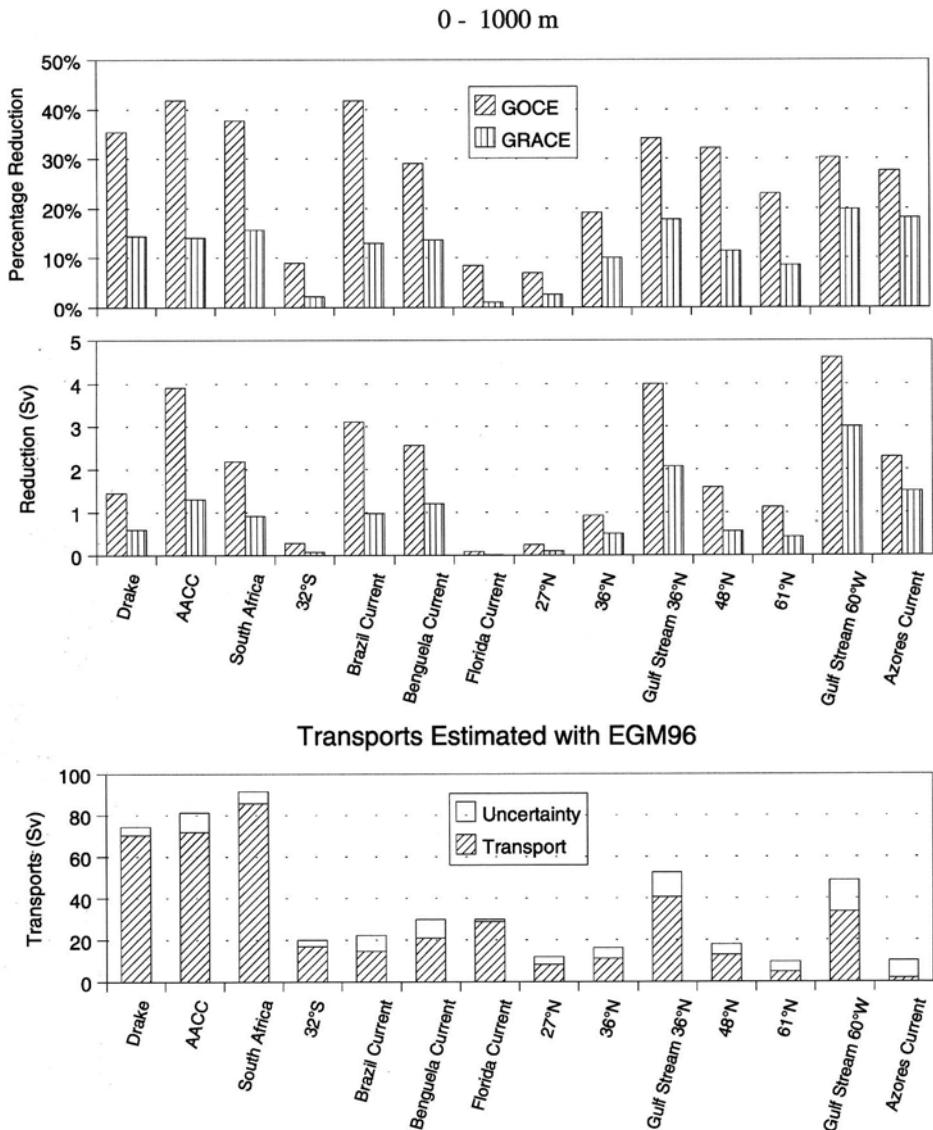


Figure 2. Impact of the GOCE and GRACE gravity missions on volume transports in the upper kilometer of the ocean. (top) Percentage reduction of transport uncertainties relative to EGM96 reference uncertainties. (middle) Corresponding absolute reduction in Sverdrups. (bottom) Transports and associated uncertainties found in the reference EGM96 calculation.

in the example of section 2, using a fully non-linear inverse model constrained by pseudo in situ observations generated by a general circulation model of the ocean circulation. In this study, uncertainties in the temperature and in the salinity fields are explicitly considered and all the contributions to the uncertainties in heat fluxes are taken into account. The results obtained by Schröter *et al.* (2002) confirm the results obtained in previous studies with a reduction of heat flux uncertainties of 37% for GOCE and 29% for GRACE in the Gulf Stream area (Table I). The pseudo in situ observations used in this study introduce the idea of synoptic high quality observations of the ocean. The following section will show that obtaining this kind of observations in the real ocean will become possible in the near future.

5. Complementarities of Space Gravity Missions and Other Observing Techniques

Ideally, space gravity observations and other observations should all be obtained simultaneously for consistency. Because the marine geoid height evolves slowly, it can be considered to be constant relative to the fast pace at which other quantities evolve. Thus, once the geoid has been estimated, it can be used for many years in order to derive the absolute dynamic topography of the ocean from altimetric observations. However, the previous sections showed that altimetric observations of the surface of the ocean must be combined with in situ observations in order to estimate the ocean circulation at all depths. The problem is that in situ observations have so far been difficult and expensive to obtain. Oceanographers have thus relied on climatologies that average hydrographic observations over many years in order to get a reasonably good spatial coverage. This averaging, unfortunately, prevents the estimation of the instant ocean circulation. Until recently, all that could be obtained at basin scale was an estimate of the “mean ocean circulation” over the many years spanned by the climatology.

This situation is about to change with the advent of ocean in situ observing networks. These networks, inspired by the network of sounding balloons used in meteorology, consist in a large number of profiling floats that drift in the ocean at a parking depth that can be as deep as 2000 m. At the end of pre-programmed periods, which range from a few days to a few weeks, the floats go back up to the surface of the ocean. While doing so, they sample vertical profiles of temperature and salinity. Once surfaced, the floats transmit the data to a ground station through a satellite connection. The float network can thus sample the interior of the ocean at regular intervals, just like altimetric missions sample the surface of the ocean every 10 to 30 days. When a complete network of floats is deployed (Figure 3), oceanographers will be able to map the hydrographic properties of the ocean in a quasi-synoptic way. Such a network is currently being deployed in the framework of the international ARGO program (<http://www-argo.ucsd.edu>; Le Traon, this volume) and should be available before the launch of the GOCE mission.

The hydrographic fields, just like traditional shipborne data, will provide velocity estimates relative to a reference level velocity.* Space gravity missions, by allowing the interpretation of altimetric observations in terms of absolute surface velocities will then provide the complementary information on reference level velocities. The uncertainties in the density field will be reduced because the noise introduced in the climatological estimates by the ocean natural variability will no longer be present in the quasi-instantaneous float observations. One may thus hope that the transmission of the very precise information provided by satellite observations to the deep layers of the ocean will no longer be hampered by the uncertainties in the density field. The challenge, however, will be to deploy an observing network that has a resolution sufficient to resolve the spatial scales of the large-scale ocean circulation. If the resolution of the network is not high enough, the data can still be assimilated in general circulation models in order to provide analyzed density fields, similar to meteorological re-analyses. The inverse approach will then serve to estimate not only the velocity field, which can be done directly by the data assimilation procedure, but also its uncertainty, which will remain out of reach of the data assimilation approach in the near future.

One final point is that the various gravity missions already undergoing or planned for the near future will also be complementary. GOCE, combined with altimetry, will provide precise estimates of the absolute ocean circulation at the surface. GRACE will provide estimates of temporal variations of bottom pressure signals at the bottom of the ocean. The difference in bottom pressure sampled by GRACE on both sides of any trans-oceanic section will provide a constraint on the temporal variations of the integrated geostrophic transport across that section. Indeed, the formulation of the geostrophic balance at depth is:

$$fu(h) = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad ; \quad fv(h) = +\frac{1}{\rho} \frac{\partial p}{\partial x} \quad (6)$$

where p denotes the pressure at depth h . Integrating the equation for the meridional velocity for instance, one obtains:

$$f \int_{x_e}^{x_w} v(h) dx = fT = \frac{1}{\rho} (p_e - p_w) \quad (7)$$

where T is the zonally integrated transport per unit depth and p_e and p_w are the bottom pressures measured by GRACE at the extremities of the section (longitudes x_e and x_w). By providing a series of bottom pressure measurements along the bathymetric line, GRACE will constrain the deep oceanic transports that are the least-constrained by GOCE.

* Float displacements during the drifting period can in principle be used to estimate absolute current velocities but the interpretation of these displacements is difficult because the amount of drift occurring while the float is at the surface of the ocean can be as large as the drift occurring while the float is at its parking depth.

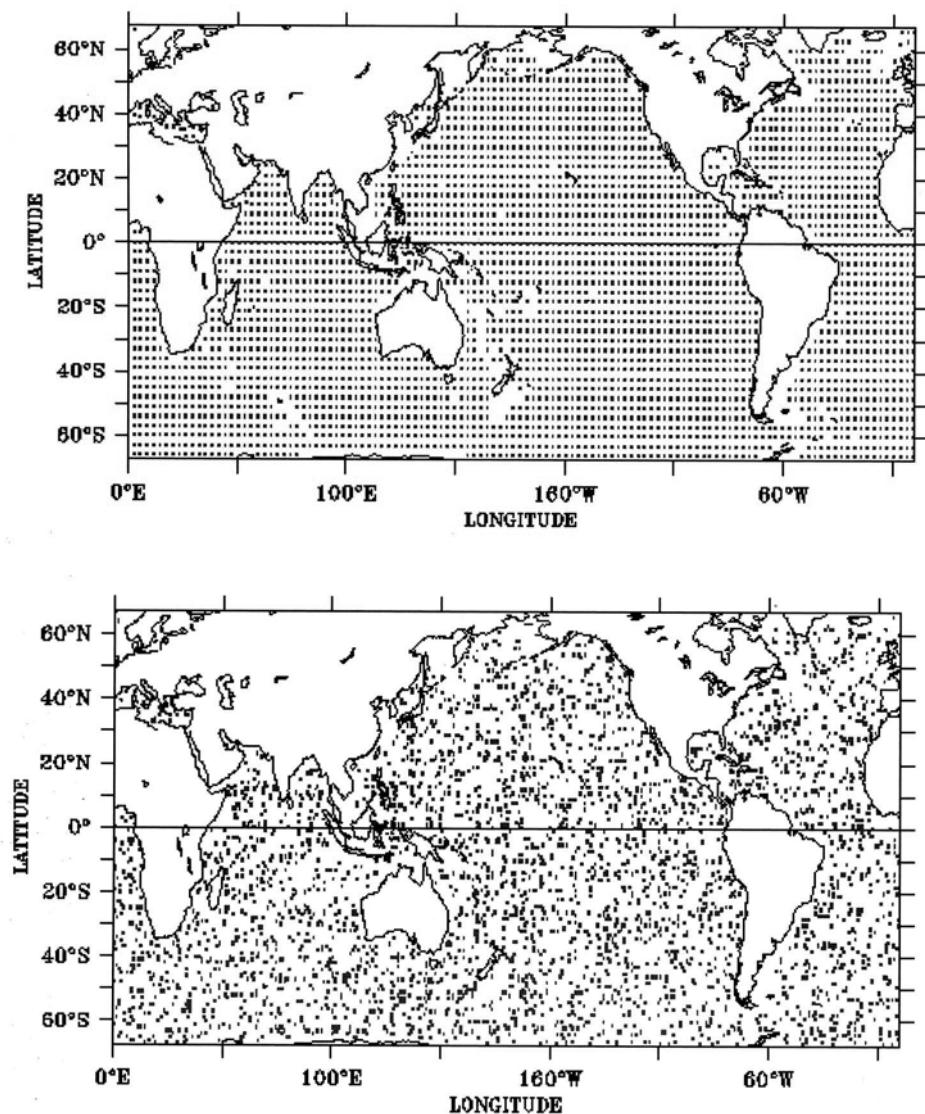


Figure 3. Initial float distribution and float distribution after some time assumed during the ARGO planning phase.

In summary, numerous issues such as the estimation of Ekman transports and the reduction in observational errors in hydrographic data through the implementation of a synoptic observing network will need to be addressed, but if successful, one may expect that space gravity observations will greatly improve our knowledge of the ocean circulation.

Acknowledgements

I wish to thank two anonymous reviewers for their help in improving the original manuscript. I gratefully acknowledge Prof. Klees and the Technical University of Delft for having hosted me for a one-year stay during which this paper was written.

References

- Coachman, L.K., and K. Aagaard: 1988, 'Transports through the Bering Strait: Annual and interannual variability', *J. Geophys. Res.* **93**, 15,535–15,539.
- Fofonoff, N.P.: 1985, 'Physical properties of seawater: new salinity scale and equation of state for seawater', *J. Geophys. Res.* **90**, 3,332–3,342.
- IPCC Third Assessment Report: 2001, 'Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)', J. T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P. J. van der Linden and D. Xiaosu (Eds.), Cambridge University Press, UK. pp 944.
- Ganachaud, A., and C. Wunsch: 2000, 'Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data', *Nature* **408**, 453–457.
- Ganachaud, A., C. Wunsch, M.-C. Kim, and B. Tapley: 1997, 'Combination of TOPEX/Poseidon data with a hydrographic inversion for determination of the oceanic general circulation and its relation to geoid accuracy', *Geophys. J. Int.* **128**, 708–722.
- Hall, M.M., and H.L. Bryden: 1982, 'Direct estimates and mechanisms of ocean heat transport', *Deep-Sea Res.* **29**, 339–359.
- LeGrand, P., H. Mercier, and T. Reynaud: 1998, 'Combining T/P altimetric data with hydrographic data to estimate the mean dynamic topography of the North Atlantic and improve the geoid', *Ann. Geophys.* **16**, 638–650.
- LeGrand, P.: 2001, 'Impact of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission on ocean circulation estimates. Volume fluxes in a climatological inverse model of the Atlantic', *J. Geophys. Res.* **106**, 19,597–19,610.
- LeGrand, P., E.J.O. Schrama, and J. Tournadre: 2002, 'An inverse modeling estimate of the dynamic topography of the ocean', *Geophys. Res. Letters*, 30(2), 34-1, CiteID 1062, DOI 10.1029/2002GLO14917
- 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, and T.R. Olson: 1998, 'The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96', : *NASA Technical Publication*, NASA Technical Publication NASA, Center for Aerospace Information. Hanover, MD.
- Milliff, R.F., M.H. Freilich, W.T. Liu, R. Atlas, and W.G. Large: 2001, 'Global ocean surface vector wind observations from space', Observing the Oceans in the 21st Century, p102-119, C.J. Koblinsky and N.R. Smith (Eds.), GODAE Project Office and Bureau of Meteorology, Melbourne.
- Pond, S., and G.L. Pickard: 1983, 'Introductory dynamical oceanography', 329 pp., Pergamon Press, Oxford.
- Schröter, J., M. Losch, and B. Sloyan: 2002, 'Impact of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission on ocean circulation estimates 2. Volume and heat fluxes across hydrographic sections of unequally spaced stations', *J. Geophys. Res.* **107**, 4-1–4-20.
- Sigman, D.M., and E.A. Boyle: 2000, 'Glacial/interglacial variations in atmospheric carbon dioxide', *Nature* **407**, 171–174.
- Wunsch, C.: 1978, 'The North Atlantic general circulation west of 50W determined by inverse methods', *Rev. Geophys. and Space Phys.* **16**, 583–620.

- Wunsch, C.: 1984, 'An eclectic Atlantic Ocean circulation model. Part I: The meridional flux of heat', *J. Phys. Oc.* **15**, 1521–1531.
- Wunsch, C.: 1996, 'The Ocean Circulation Inverse Problem', 422 pp., Cambridge Univ. Press, New York.

Address for Offprints: P. LeGrand, Laboratoire de Physique des Oceans, IFREMER, BP 70, F-29280 Plouzane, France, plegrand@ifremer.fr

HOW OPERATIONAL OCEANOGRAPHY CAN BENEFIT FROM DYNAMIC TOPOGRAPHY ESTIMATES AS DERIVED FROM ALTIMETRY AND IMPROVED GEOID

P.Y. LE TRAON, F. HERNANDEZ, M.H. RIO and F. DAVIDSON

CLS Space Oceanography Division, 31526 Ramonville st Agne, France

(pierre-yves.lettraon@cls.fr)

Received: 7 August 2002; Accepted in final form: 6 November 2002

Abstract. With a precise geoid, GOCE will allow an estimation of absolute dynamic topography from altimetry. The projected benefits to operational oceanography and its applications are analyzed herein. After a brief overview of operational oceanography, we explain how the new geoids will be used in the future to improve real time altimeter products and to better constrain modelling and data assimilation systems. A significant impact is expected both for mesoscale (*e.g.* better estimations and forecasts of currents for pollution monitoring, marine safety, offshore industry) and climate (better initialization of coupled ocean/atmosphere models) applications.

1. Introduction

A main objective of the ESA GOCE mission is to provide a geoid of sufficient accuracy to allow precise estimation of absolute dynamic topography (*i.e.* sea level relative to the geoid) from altimetry. This will have a major contribution to oceanography (see Woodworth *et al.*, 1998, Le Provost *et al.*, 1999). The projected benefits to operational oceanography and its applications are analyzed herein. The paper is organized as follows. An overview of operational oceanography and of the Global Ocean Data Assimilation Experiment (GODAE) is given in section 2. We explain, in section 3, how altimeter data are processed and used for near real time applications. Section 4 describes current approaches to get absolute dynamic topography from altimetry while the impact of new geoids for operational oceanography is analyzed in section 5. We conclude the paper by highlighting future research directions in the context of the GOCE mission.

2. Operational Oceanography and GODAE

A significant proportion of world economic and social activities depends on the sea. These activities are subject to loss of efficiency and direct costs and damages caused by the varying impact of the marine environment. To mitigate against such effects, an accurate description and a reliable prediction of the ocean variations are crucial. The Global Ocean Data Assimilation Experiment (GODAE) belief is



Space Science Reviews **108:** 239–249, 2003.

© 2003 Kluwer Academic Publishers.

that this will be best achieved through a global and integrated approach combining model, *in situ* and remote sensing data through data assimilation. The vision is “A global system of observations, communications, modeling and assimilation, that will deliver regular, comprehensive information on the state of the oceans, in a way that will promote and engender wide utility and availability of this resource for maximum benefit to the community” (The International GODAE Steering Team, 2001).

GODAE will conduct its main demonstration phase from 2003 to 2005. From 2003 to 2005, operational and research institutions from Australia, Japan, United States, United Kingdom, France and Norway will perform global oceanic data assimilation and ocean forecasts to provide regular and comprehensive descriptions of ocean state variables at high temporal and spatial resolution. This demonstration phase will be followed by a consolidation and transition phase spanning years 2005 to 2007 where the synthesis and transition to operational systems will take place.

Le Traon *et al.* (2001) provide a review of the main operational applications that will benefit from the GODAE integrated approach; they include climate and seasonal forecasting, navy applications, marine safety, fisheries, the offshore industry and management of shelf/coastal areas. For a more general discussion of the benefits and users of operational oceanography, the reader is referred to Flemming (1997), IOC (1998) or Flemming (2001).

3. Satellite Altimetry and Operational Oceanography

3.1. CONTRIBUTION OF SATELLITE ALTIMETRY

Satellite altimetry is probably the most important remote sensing data set currently available for operational oceanography. It has unique capabilities for providing a global high resolution synoptic view of ocean circulation. Altimetry provides measurements of sea surface topography (sea level) which includes depth integrated oceanic information. These measurements provide a strong constraint for 3D ocean circulation estimation. Furthermore satellite altimetry is a mature technique and data can now be processed in near real time for operational applications.

Operational oceanography and GODAE require at least two (and preferably three) simultaneous altimeter missions including one highly accurate long term altimeter system. The latter is mandatory for climate applications and as a reference for the other altimeter missions. The combination of data issued from several altimeters is needed to constrain mesoscale circulation and provide useful surface velocity field information, a requirement for most operational activities.

3.2. THE SSALTO/DUACS REAL TIME PROCESSING SYSTEM

The CLS/CNES SSALTO/DUACS system is a near real time multiple altimeter processing system. It has been designed to meet both climate and mesoscale applications. It is able to acquire, process and merge TOPEX/POSEIDON (TP), ERS-2, GEOSAT Follow On (GFO), Jason-1 and ENVISAT data in near real time. The most up-to-date geophysical corrections are applied along with improved orbit error and long wavelength error reduction schemes. Near real time products include along-track Sea Level Anomaly (SLA) for the different satellites and a high resolution SLA map merging the different satellites.

3.3. USE OF ALTIMETER DATA FOR OPERATIONAL OCEANOGRAPHY

Altimeter data are now used for a wide range of operational oceanography applications (*e.g.* Digby *et al.*, 1999). SSALTO/DUACS products are thus used to monitor eddies and currents as well as for providing long-term statistics on surface currents for offshore applications. They have also been successfully tested by fishing fleets to help locate favorable fishing grounds. The same information can be used by national agencies in charge of managing fish stocks to help them better assess these stocks and understand how they are impacted by changes of the oceanic environment.

Most applications rely, however, on data assimilation systems that allow an effective integration of remote sensing data with *in situ* data and models to provide analyses and forecasts of the ocean state. SSALTO/DUACS is thus presently serving GODAE centers (including MERCATOR, the French component to GODAE), the French Navy SOAP system and seasonal and climate forecasting centers such as ECMWF. All these forecasting centers serve mesoscale (*e.g.* marine safety, oil pollution monitoring, boundary conditions for coastal/shelf models, Navy acoustic applications) and/or climate applications (*e.g.* seasonal predictions).

4. Absolute Dynamic Topography

Satellite altimetry provides Sea Level Anomaly ($\eta' = \eta - \langle \eta \rangle$) with good accuracy (a few cm) and high resolution. In practice the best procedure to get an estimation of the absolute dynamic topography ($\eta = \langle \eta \rangle + \eta'$) from altimetry will be:

- Estimate a mean dynamic topography (MDT, $\langle \eta \rangle$). The mean should correspond to a mean over a precise time period determined by the sea level anomaly calculation (typically several years or the duration of an altimetric mission). A longer averaging time is preferred as the resulting mean signal spatial scales will be larger.
- Add this MDT to sea level anomaly (η') derived from repeat-track analysis.

4.1. MEAN DYNAMIC TOPOGRAPHIES

The mean dynamic topography $\langle \eta \rangle$ can be obtained from:

1. The difference between an altimeter Mean Sea Surface (MSS) and a geoid,
2. from *in situ* data,
3. from a model,
4. from the combination of 1/, 2/, 3/ through data assimilation or inverse modeling.

To get a precise estimation of instantaneous absolute dynamic topography, $\langle \eta \rangle$ must be known with a resolution of 100–200 km and an accuracy of a few cm (e.g. Le Provost *et al.*, 1999, and Le Provost and Bremond, 2003). It must also correspond to a specified time period.

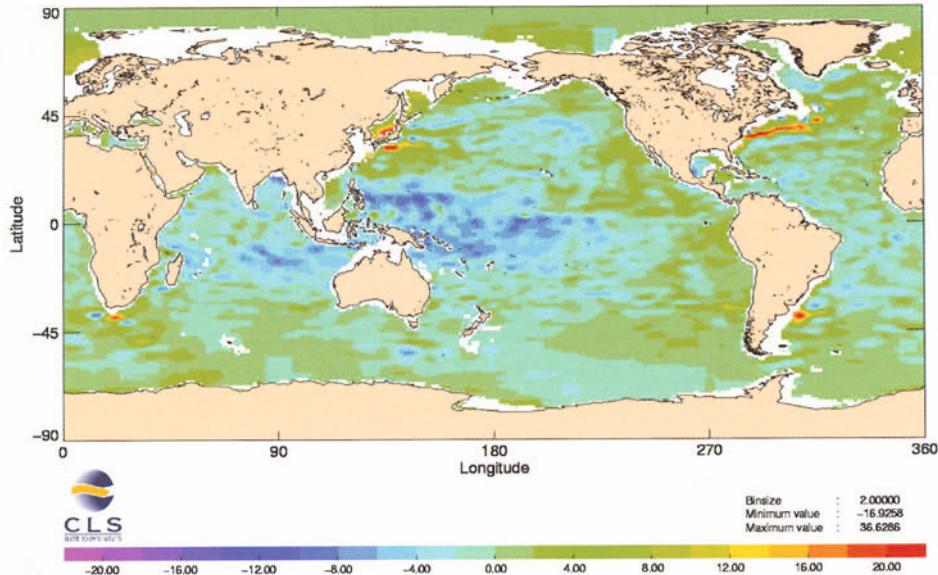
Geoid models currently available to the oceanographic community are not sufficiently accurate to provide a useful estimation of mean dynamic topography. Typical accuracy is of 10 to 20 cm rms for a spherical harmonic development of the order of 20 (*i.e.* wavelengths larger than 2000 km) (e.g. Lemoine *et al.*, 1998). This is of the order or larger than the existing ocean models or *in situ* data derived MDT accuracy. In addition these geoids are not independent from altimetric data and have absorbed part of the oceanic signal. In particular, they contain the oceanic signal that is above the cut-off dynamic topography expansion degree used in the geoid model representation.

MDTs have also been derived in the past from climatological *in situ* data. The most widely used mean dynamic topography is the one derived from the Levitus climatology. These estimations require to assume a reference level of no motion and thus miss out part of the surface dynamic topography. This is true particularly at high latitudes where the barotropic component of the flow can be large (e.g. Antarctic Circumpolar Current). Deep floats can be used, however, to provide an estimation of the missing signal at the reference level (e.g. Davis, 1998). This can be done directly or by adding dynamical constraints through an inverse method (e.g. Legrand and Minster, 1999). One of the disadvantages of such estimations, however, is that to reduce the ocean variability noise, they require *in situ* data spanning several decades. The resulting MDTs thus do not correspond to the averaging time period compatible with the altimeter sea level anomaly data calculation.

4.2. SSALTO/DUACS AND MERCATOR APPROACH

The MDT estimation undertaken by SSALTO/DUACS to serve MERCATOR needs is outlined here. This work is also a component of the EC ENACT project aimed at developing improved altimeter and *in situ* data assimilation methods for climate and seasonal forecasting.

Altimeter sea level anomalies and simultaneous *in situ* data are used to estimate a “synthetic geoid” or more exactly a “synthetic mean dynamic topography”. The technique proceeds as follows. *In situ* data can provide estimates of the absolute



Units are cm. This shows the expected errors when Levitus climatology is used to reference altimeter data. Errors are large (> 10 cm) in the tropical Pacific ocean because of interannual variability and in western boundary currents that require a better resolution (from Guinehut, 2002).

dynamic topography η (although the barotropic part maybe more difficult to estimate) and satellite altimetry provides η' at the position and time of *in situ* data. The combination of the two estimates thus yields the mean dynamic topography $\langle \eta \rangle$ over the needed time period. This methodology while powerful requires a large amount of simultaneous data. This technique has been applied in the Gulf Stream (Mitchell *et al.*, 1990), the Kuroshio (Imawaki and Uchida, 1997) and the Azores Current (Hernandez, 1998).

The technique has been applied for the first time to the global ocean using *in situ* temperature and salinity profile data (XBT, CTD) and T/P and ERS-1/2 altimeter data (Guinehut, 2002). This allows tuning of the Levitus mean dynamic topography (here dynamic height relative to 700 m) so that it corresponds to an improved mean dynamic topography over a specified time period. The improvement is particularly noticeable in western boundary currents with sharper MDT fronts and in the tropical Pacific ocean where the new MDT differs by up to 10 cm from the Levitus only MDT (Figure 1). This latter discrepancy stems from the large inter-annual variability in the tropical Pacific which is not seen by the Levitus Climatology, yet would influence the MDT over the several years spanning an altimeter mission.

This new mean dynamic topography enhances altimeter versus *in situ* data agreement. For example, the rms difference between a global XBT data set and T/P-ERS altimeter data is reduced from 9.5 cm (using MDT from Levitus Climatology) to 7.5 cm rms using the new mean dynamic topography.

A similar calculation has been done with surface drifters (see also Imawaki *et al.*, 2003). Velocities were corrected first from wind effects using an empirical Ekman model. They were then corrected for ocean variability using T/P and ERS-1/2 and a mean dynamic topography was computed. This mean dynamic topography also provides an improved comparison with *in situ* data. Work is underway to merge the drifter, temperature/salinity estimations with a mean dynamic topography derived from CHAMP geoid (and possibly GRACE). Prior to the combination, a barotropic component (more exactly here the dynamic topography signal at 700 meters) must be added to the temperature and salinity estimates. Use of deep floats (*e.g.* Davis, 1998) and/or model outputs are envisioned for this purpose.

This method already provides a better estimation of the MDT but is limited by the number of *in situ* measurements and cannot be applied in coastal/shelf regions. Its precision is also low in high latitude regions where the barotropic signals are large. This will improve in the future with the Argo array of profiling floats (Roemmich *et al.*, 2001) that is considerably increasing the available number of temperature/salinity data. The Argo data are also providing estimate of the velocity field at depth that could be used to derive the barotropic component. Note that such a calculation should be best achieved in the future with effective data assimilation techniques that could merge SLA altimeter data with all available *in situ* data and high resolution models. This requires advanced multivariate data assimilation techniques that are not yet available but should be developed as part of GODAE.

5. Impact for Operational Oceanography and Ocean Forecasting

While a significant progress can be made in the coming years to improve MDT estimations, only GOCE will be able to provide the 1 or 2 cm precision with a 100 or 200 km resolution everywhere that is needed to adequately resolve the MDT signals. This is expected to have a large impact for operational oceanography applications.

5.1. IMPACT FOR OPERATIONAL ALTIMETRY PRODUCTS

With precise MDT estimates, surface currents will be much better known in regions such as western boundary currents where the mean kinetic energy is comparable to the eddy kinetic energy. This will have a direct application for offshore, pollution monitoring and marine safety applications. Knowing precisely the MDT will also be very useful to unambiguously interpret altimeter data (Figure 2). Today, it is often difficult to interpret anomaly fields (sea level or velocity) and make the difference, for example, between a meander or an eddy. This limits the use of altimeter data both for scientific applications and operational applications.

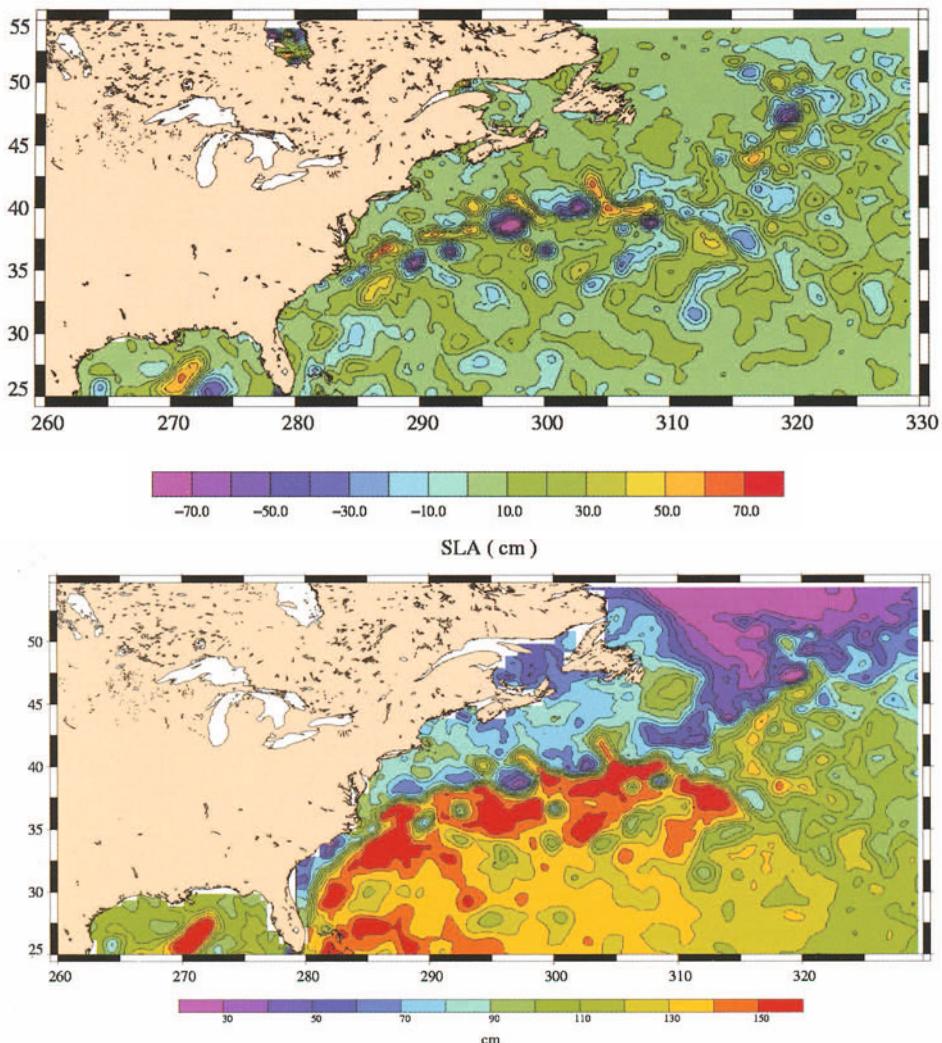


Figure 2. Map of SLA from TOPEX/POSEIDON and ERS-2 data in the Gulf Stream region (upper figure). Units are cm. Corresponding absolute dynamic topography obtained using a synthetic mean dynamic topography (lower figure). The absolute topography provides a better localization of the main frontal areas and their eddies or meanders and allow an unambiguous interpretation of altimeter signals. This is crucial for operational oceanography applications.

5.2. IMPACT FOR OCEAN STATE ESTIMATION

Most modeling/assimilation systems use an external mean dynamic topography (MDT) to get absolute dynamic topography. In practice, the MDT is often the model mean resulting in an (basically) uncorrected model mean (*e.g.* Killworth *et al.*, 2001). In addition, when altimeter data are combined with *in situ* data (*e.g.*

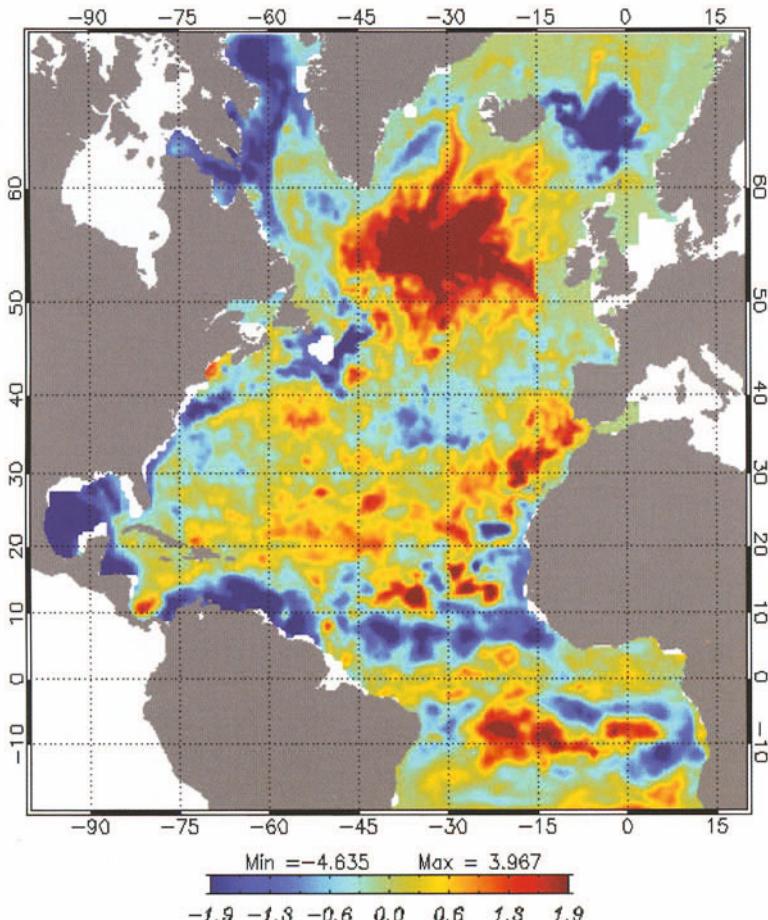


Figure 3. Difference in the mean temperature field at 100 m of the MERCATOR data assimilation system when the operational MDT is replaced by a synthetic MDT. This illustrates the impact of MDT errors on ocean state estimation. Units are °C.

multivariate data assimilation), MDT error is the main source of inconsistency. Guinehut (2002) thus showed that when altimeter data are compared with *in situ* data, the MDT is the largest source of error; the model MDT must thus be improved if one wants to effectively combine altimeter and *in situ* data.

The assimilation of a model derived MDT from one model into another, constrains, however, the mean circulation such that both models resemble each other more (Willebrand *et al.*, 2001). This suggests that with a correct MDT we expect the ocean model circulation to further approach that of the real ocean. This is encouraging in the event when we can derive an accurate MDT from the combination of precise altimetry and geoid data.

With a precise MDT, we can thus expect a significant impact on ocean state estimation and its mesoscale and climate applications. Seasonal to interannual

forecasts are, in particular, heavily dependant on the ocean initial state estimation. Today, because of the uncertainties of MDT, assimilation of altimeter data alone can induce significant biases in the ocean state estimation. As an illustration, Figure 3 shows the difference in the mean temperature field at 100 meters estimated from the MERCATOR system with two different MDTs; the first MDT is the one used in the operational system. It corresponds to the model mean with a correction to reduce model and satellite-observed sea-surface temperature mismatch. The second model is from a synthetic climatology estimation (see section 4.2). Differences can reach several degrees and illustrate how today errors in the MDTs impact the ocean state estimation.

5.3. IMPACT FOR SHORT-TERM OCEAN FORECASTING

Improved MDTs from geoid and altimetry should also allow a better forecast of the ocean circulation. Upcoming and current gravity missions when combined with altimetry will improve absolute dynamic topography precision. This will allow better ocean model initialization and improved prediction skill. For example, if a MDT positions the Gulf Stream differently from reality, the dynamics of the model mean currents would be inconsistent with the dynamics of turbulence associated with the assimilation of SLA data. In this case the assimilation would push the model towards a dynamically inconsistent state, that would be short lived once assimilation is removed for prediction purposes. To quantify this effect, simulations were performed with the SHOM SOPRANE forecasting system as part of an ESA GOCE study (Le Provost *et al.*, 1999). Simulated altimeter data and mean dynamic topography were derived from the Los Alamos 1/10° model (Smith *et al.*, 2000). Two tests were performed:

1. Use of MDT resolving scales longer than 200 km (“GOCE” MDT).
2. Use of a MDT resolving scales longer than 2000 km only (the “today” case).

We have shown that the use of the future “GOCE” MDT would allow better prediction of the eddy field, by roughly 15% of the signal variance for a 7-day prediction, and that 2-week forecast error using the “GOCE” MDT is equivalent to the 1-week forecast error with the “today” MDT. This improvement will come in addition to the improvement in the knowledge of the mean circulation itself. The effect should be larger for a longer term forecast.

6. Conclusions — Prospects for GOCE

GOCE provides a unique opportunity to solve the “mean problem” in altimeter data assimilation into ocean models. GOCE will provide a substantial and timely contribution to operational oceanography by improving analyses and forecast quality of ocean state (eddy resolving models).

To benefit from the full potential of this mission, work should be carried out:

- to define the best strategy to estimate a mean dynamic topography (MDT) *and its error* from GOCE (GRACE/CHAMP) and altimeter data (mean sea surface).
- to define the best methodologies to validate such MDTs and their errors (*and corresponding ocean circulation*). The comparison with MDTs derived from *in situ* data and models (e.g. through synthetic method, inverse modeling or data assimilation) will be very useful for the validation. With the development of new *in situ* observing techniques such as Argo, these MDTs could reach, indeed, an accuracy of a few cm rms for scales larger than a few hundreds of km. Finally, the comparison of absolute dynamic topography derived from GOCE MDTs and altimeter SLA with *in situ* data (drifters, Argo) should also be very powerful.
- to combine these MDTs with those derived from *in situ* measurements. The *in situ* and GOCE/GRACE/CHAMP topographies will have very different error characteristics and will thus be complementary. It is likely that the *in situ* data estimations will be very accurate in well sampled regions and low variability regions. In other regions and also at high latitudes and in coastal regions (where the barotropic component will not be well estimated) they will be less useful.
- to quantify the impact of these new MDTs in a global data assimilation system. This should be done both for analyses and forecasts.

References

- Davis, R.E.: 1998, 'Preliminary results from directly measuring mid-depth circulation in the tropical and South Pacific', *J. Geophys. Res.* **103**, 24619–24639.
- Digby, S., Antczak, T., Leben, R., Born, G., Barth, S., Cheney, R., Foley, D., Goni, G. J., Jacobs, G., and Shay, N. : 1999, 'Altimeter data for operational use in the marine environment', *IEEE/MTS Conference Oceans 99*, Seattle, September 1999, 9 pp.
- Flemming, N.C.: 1997, 'Estimates of the costs and benefits of operational oceanography at the single industry level', in J. Stel (ed.), *Operational Oceanography, the challenge for European co-operation*, Elsevier Oceanography Series **62**, pp. 269–277.
- Flemming, N. C.: 2001, 'Dividends from Investing in Ocean Observations: A European Perspective', in C.J. Koblinsky and N.R. Smith (eds.), *Observing the Oceans in the 21st Century*, GODAE Project Office, Bureau of Meteorology, pp. 66–84.
- IOC (Intergovernmental Oceanographic Commission): 1998, *The GOOS 1998*, GOOS Publication No. 42, Paris, pp. 144.
- Guinehut, S.: 2002, 'Vers une utilisation combinée des données altimétriques et des mesures des flotteurs profilants', Phd Thesis, Université Paul Sabatier, Toulouse, 168 pp.
- Hernandez, F.: 1998, 'Estimation d'un géoïde synthétique sur la zone Semaphore', Rapport final. Contrat SHOM/CLS No. 97.87.003.00.470.29.45, 32 pp.
- Imawaki, S. and Uchida, H.: 1997, 'Time series of the Kuroshio transport derived from field observations and altimetry data', *International WOCE Newsletter* **25**, 15–18.
- Imawaki, S., Uchida, H., Ichikawa, K., and Ambe, D.: 2003, 'Estimating the high-resolution mean sea-surface velocity field by combined use of altimeter and drifter data for geoid model improvement', *Space Sci. Rev.*, this volume.

- Killworth, P. D., Dieterich, C., Le Provost, C., Oschlies, A., and Willebrand, J.: 2001, 'Assimilation of altimetric data and mean sea surface into an eddy-permitting model of the North Atlantic', *Progress in Oceanography* **48**, 313–335.
- Le Grand, P. and Minster, J. F.: 1999, 'Impact of the GOCE gravity mission on ocean circulation estimates', *Geophys. Res. Letters* **26**, 1881–1884.
- Lemoine, F. G. et al.: 1998, *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 20771 USA, 466 pp.
- Le Provost, C., Le Grand, P., Dombrowsky, E., Le Traon, P. Y., Losch, M., Ponchaud, F., Schröter, J., Sloyan, B., Sneeuw, N.: 1999, *Impact of GOCE for ocean circulation studies*, ESA study contract 13175/98/NL/GD, Final Report, 157 pp.
- Le Provost, C., and Bremond, M.: 2003, 'Resolution needed for an adequate determination of the mean ocean circulation from altimetry and improved geoid', *Space Sci. Rev.*, this volume.
- Le Traon, P.-Y., Rienecker, M., Smith, N., Bahurel, P., Bell, M., Hurlburt, H., and Dandin, P.: 2001, 'Operational Oceanography and Prediction — a GODAE Perspective', in C. J. Koblinsky and N. R. Smith (eds.), *Observing the Oceans in the 21st Century*, GODAE Project Office, Bureau of Meteorology, pp. 529–545.
- Mitchell, J. L., Dastugue, J. M., Teague, W. J., Hallock, Z. R.: 1990, 'The estimation of geoid profiles in the northwest Atlantic from simultaneous satellite altimetry and airborne expendable bathythermograph sections', *J. Geophys. Res.* **95**, 17,965–17,977.
- Roemmich, D., Boebel, D. O., Desaubies, Y., Freeland, H., Kim, K., King, B., Le Traon, P. Y., Molinari, B., Owens, B., Riser, S., Send, U., Takeuchi, K., and Wijffels, S.: 2001, 'Argo: The Global Array of Profiling Floats', in C. J. Koblinsky and N. R. Smith (eds.), *Observing the Oceans in the 21st Century*, GODAE Project Office, Bureau of Meteorology, pp. 248–258.
- Smith, R. D., Maltrud, M. E., Bryan, F. O., and Hecht, M. W.: 2000, 'Numerical simulation of the North Atlantic ocean at 1/10°', *J. Phys. Oceanogr.* **30**, 7,153–21561.
- The International GODAE Steering Team: 2001, *The Global Ocean Data Assimilation Experiment Strategic Plan, GODAE report no. 6*, GODAE International Project Office, Australia, Melbourne, 26 pp.
- Willebrand, J., Barnier, B., Böning, C., Dieterich, C., Killworth, P. D., Le Provost, C., Jia, Y., Molinari, J. M., New, A. L.: 2001, 'Circulation characteristics in three eddy-permitting models of the North Atlantic', *Progress in Oceanography* **48**, 123–161.
- Woodworth, P. L., Johannessen, J., Le Grand, P., Le Provost, C., Balmino, G., Rummel, R., Sabadini, R., Suenkel, H., Tscherning, C. C., Visser, P.: 1998, 'Towards the definitive space gravity mission', *International WOCE Newsletter* **33**, 37–40 and 24.

IV: GEODESY

REMARKS ON THE ROLE OF HEIGHT DATUM IN ALTIMETRY-GRAVIMETRY BOUNDARY-VALUE PROBLEMS

FAUSTO SACERDOTE

Dipartimento di Ingegneria Civile, Università di Firenze, Italy (fausto@dicea.unifi.it)

FERNANDO SANSÒ

DIIAR - Facoltà di Ingegneria di Como, Politecnico di Milano, Italy (fsanso@geo.polimi.it)

Received: 18 March 2002; Accepted in final form: 26 September 2002

Abstract. The problem of global geoid determination is usually solved using satellite altimetry data on the oceans, together with an oceanographic model of sea surface topography, and gravity anomaly data on the continents. Such data, however, enable to obtain only potential differences with respect to a reference surface whose absolute potential is unknown. This situation suggests to modify the classical mixed boundary-value problem of physical geodesy by inserting into the boundary conditions an unknown additive constant, that must be determined by imposing a suitable additional constraint. Yet, such formulation of the boundary-value problem, from the point of view of its mathematical properties, is not unconditionally well-posed, and, furthermore, does not reflect faithfully the available physical model, as the present knowledge of ocean circulation does not allow to connect along coastlines the reference surfaces defined on the oceans and on the continents. The introduction of two different unknown additive constants, one for the oceans and one for the earth, to be determined by imposing two additional constraints, gives rise to a more faithful picture of the present physical knowledge, and, at the same time, to a new well-posed formulation of the boundary-value problem.

Keywords: height datum, geoid, boundary-value problems

1. Introduction

In most of the classical boundary-value problems of physical geodesy it is assumed that the value of the gravity potential is known on the whole boundary surface or at least on part of it. Actually, using gravity measurements and levelling, only potential differences can be obtained, whereas the absolute value of the geopotential cannot be obtained at any point with acceptable accuracy. Consequently, the boundary values of the geopotential must be assumed to be known except for one additive constant, as remarked for example by Svensson (1991) for non-linear Molodensky problem. According to Svensson, the value of this constant is fixed if GM (the gravitational constant multiplied by the mass of the earth) is assumed to be known, *i.e.* it is given as an additional information. Consequently, when the problem is linearized, introducing a reference potential, the reference surface cannot be defined assuming the reference potential on it to be exactly equal to the geopotential on the earth's surface, but a small constant gap must be introduced. The determination of



Space Science Reviews **108:** 253–260, 2003.

© 2003 Kluwer Academic Publishers.

this constant is particularly simple for Stokes's problem (spherical approximation, spherical boundary — see Heiskanen and Moritz, 1967, section 2–19):

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \text{ (exterior of } S) \\ -T' - \frac{2}{R}T + \frac{2}{R}\delta W = \widetilde{\Delta}g & \text{on } S \\ T = 0(r^{-3}) \end{cases} \quad (1)$$

For the sake of simplicity, it is assumed that $\widetilde{\Delta}g$ does not contain first degree harmonic components, so that the problem is solvable without further modifications. Owing to the asymptotic condition, it is required that the solution T does not contain the zero-degree harmonic term; consequently, the same must be true for $\widetilde{\Delta}g - \frac{2}{R}\delta W$ in the boundary condition. Hence

$$\frac{1}{4\pi} \int_{S_1} \widetilde{\Delta}g d\sigma = \frac{2}{R}\delta W \quad (2)$$

determines uniquely the value of δW . It is well-known that, in order to obtain a disturbing potential T with vanishing zero-degree harmonic coefficient, it is necessary to assume the mass of the level ellipsoid to be equal to the mass of the earth. More generally, once the earth's mass is known, in order to define a level ellipsoid and its gravity potential, it is necessary to fix the mass of the reference normal potential too; the mass difference between earth and ellipsoid determines the zero-degree harmonic coefficient of the disturbing potential, and consequently the value of δW .

Actually, modern methods for geoid computation based on gravimetric data make use of a global geopotential model, which is determined from satellite altimetry data on the oceans and gravity anomalies on the continents; therefore, within the frame of boundary-value problem theory, the geoid determination problem is today more suitably classified as a mixed altimetry-gravimetry problem.

Geodetic mixed boundary-value problems have been extensively studied in the 1980s. In particular, for the so called Problem 1, which may be expressed, in spherical approximation, as:

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \\ T = f_1 & \text{on } S_S \text{ (sea)} \\ -\frac{\partial T}{\partial r} - \frac{2}{r}T = f_2 & \text{on } S_L \text{ (land)} \\ T = 0(r^{-1}) \end{cases} \quad (3)$$

Svensson (1985) proved analytically that, for particular distributions of the continents, the solution, defined in a suitable function space, is not unique. This fact is confirmed by a numerical analysis conducted with a simple model of a continental area covering a spherical cap with varying radius; a radius was found for which uniqueness failed. Sacerdote and Sansò (1987) showed that it is possible to obtain uniqueness introducing an unknown constant in the boundary condition on the sea, and adding a constraint on the boundary values of T , namely $\int_S T d\sigma = 0$.

Such a modified formulation was physically justified by assuming the altimetric measurements to be affected by a constant bias.

This approach, which was suggested more by the mathematical properties of the solutions of the mixed boundary-value problem (BVP) rather than by realistic physical considerations, has been subsequently criticized by Lehmann (2000), who observed that (in particular in recent years) the calibration problem of satellite altimetry was believed to be reduced to the level of very few centimeters. He was rather interested to go back to the original formulation, where the absolute value of the potential was not exactly known, but rather had to be considered as an unknown parameter, for the knowledge of which the further information on GM had to be introduced. Assuming that using suitable ocean circulation models one could connect the reference surfaces for land and ocean at tide gauges, the corresponding mixed BVP could be derived via linearization, yielding

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \\ T = f_1 + \delta W & \text{on } S_S \\ -\frac{\partial T}{\partial r} - \frac{2}{r} T = f_2 - \frac{2}{r} \delta W & \text{on } S_L \\ T = 0(r^{-2}) \end{cases} \quad (4)$$

The last condition in (4), which is stronger than the corresponding one in (3), can be equivalently written as $T_{00} = 0$, expressing the equality of the masses generating the normal and the actual gravity potential; the introduction of this new constraint balances the presence of the additional unknown δW in (4).

Lehmann (2000) showed that the problem (4) may not have a unique solution for certain distributions of the continental surface. More precisely, he studied the problem with numerical methods for an axially symmetric configuration of the boundary, and found that (similarly to the case discussed by L. Svensson) a singularity occurs in the case the continents cover a spherical cap with a certain radius.

So far the problem might appear insoluble, particularly because the assumption of knowing a realistic circulation pattern of the ocean and of having a good mathematical model to translate it into a field of height anomalies could be considered as a pure intellectual speculation. In the present paper first the problem (4) is explored by a simple analytical tool, showing also in this way that such a problem cannot be considered as well-posed unconditionally with respect to the shape of land and ocean. Moreover it is shown that (looking more carefully to what oceanography can provide to geodesy) not one constant but two have to be introduced into the BVP to account for the actual rank deficiency of the observables. For such a new formulation of the problem, suitably constrained by two conditions, a new theorem of existence and uniqueness of the solution is proved, irrespective of the shapes of land and ocean domains.

2. The Non-Unconditional Uniqueness of Altimetry-Gravimetry BVP 1 With Unknown Reference Potential

In this section it is proved analytically that the problem (4) with a spherical boundary is not well-posed for some configuration of land and oceans. The check of well-posedness relies on the application of the Fredholm alternative and on the investigation of the uniqueness of the corresponding homogeneous problem:

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \\ T = \delta W & \text{on } S_S \\ -T'_r - 2T = -2\delta W & \text{on } S_L \\ T_{00} = 0 & \end{cases} \quad (5)$$

For simplicity the radius of the spherical boundary is assumed to be unity. Introducing the new unknown function $V = T - (\delta W/r)$, it follows that $V_{00} = -\delta W$; furthermore

$$(-V'_r - 2V)|_S = (-T'_r - 2T)|_S + \delta W = -\delta W = V_{00} = \frac{1}{4\pi} \int_S V d\sigma \quad (6)$$

and consequently

$$\begin{cases} \nabla^2 V = 0 & \text{in } \Omega \\ V = 0 & \text{on } S_S \\ (-V'_r - 2V)|_{S_L} - \frac{1}{4\pi} \int_S V d\sigma = 0 & \end{cases} \quad (7)$$

Now it must be checked whether Problem (7) has non-vanishing solutions, or, equivalently, whether the corresponding eigenvalue problem

$$\begin{cases} \nabla^2 V = 0 & \text{in } \Omega \\ V = 0 & \text{on } S_S \\ (-V'_r - 2V)|_{S_L} - \frac{1}{4\pi} \int_S V d\sigma = \lambda V|_{S_L} & \end{cases} \quad (8)$$

admits a zero eigenvalue. Problem (8) can be obtained from the minimization of the functional

$$\Phi = \frac{1}{4\pi} \int_{S_L} V(-V'_r - 2V)d\sigma - \left(\frac{1}{4\pi} \int_S V d\sigma \right)^2 \quad (9)$$

under the conditions $(1/4\pi) \int_S V^2 d\sigma = 1$, $V|_{S_S} = 0$, and the value of this minimum is exactly equal to λ . The first term of Φ is meaningful if $V|_S$ belongs to the space $H^{1/2}(S)$, *i.e.*, if the solution is sought in the space $H^1(\Omega)$. Φ can be expressed as

$$\Phi = \sum_{n>1} (n-1)V_{nm}^2 - V_{00}^2 = \sum_{n>1} (n-1)V_{nm}^2 - 2V_{00}^2 \quad (10)$$

It is clear from equation (10) that $\Phi \geq -2V_{00}^2 \geq -2$ (the last inequality stems from the first minimization condition); indeed, in the particular case $S_L \equiv S$ (*i.e.*, assuming the earth's surface to be entirely covered by land), the lower bound is reached by the solution $V = 1/r$.

Hence, in the limit case $S_L = S$, λ is negative. Following a procedure similar to the one adopted by (Svensson, 1985), in order to prove that λ necessarily vanishes for some configuration of land and oceans it is sufficient to show that, if the extension of land is small enough, λ is positive; in fact, if the border between land and ocean is continuously modified, the value of λ varies continuously, so that it must cross zero.

Indeed, from the inequality

$$\begin{aligned} V_{nm}^2 &= \left(\frac{1}{4\pi} \int_S V Y_{nm} d\sigma \right)^2 = \left(\frac{1}{4\pi} \int_{S_L} V Y_{nm} d\sigma \right)^2 \leq \\ &\leq \frac{1}{4\pi} \int_{S_L} V^2 d\sigma \cdot \frac{1}{4\pi} \int_{S_L} Y_{nm}^2 d\sigma = \frac{1}{4\pi} \int_{S_L} Y_{nm}^2 d\sigma \end{aligned} \quad (11)$$

one obtains (recalling that $\sum_{m=-n}^n Y_{nm}^2 = 2n + 1$)

$$\sum_{m=-n}^n V_{nm}^2 \leq (2n + 1) \frac{\mu(S_L)}{4\pi} \equiv (2n + 1)q \quad (12)$$

where $\mu(S_L)$ is the area of the land fraction. Hence $V_{00}^2 \leq q$; $V_{00}^2 + \sum_{l=1}^n V_{lm}^2 \leq 4q$ and, considering that $(1/4\pi) \int_S V^2 d\sigma = \sum V_{nm}^2 = 1$, it follows that $\sum_{n>1} V_{nm}^2 \geq 1 - 4q$. Finally from equation (10), since $n - 1 \geq 1$ when $n > 1$, one may conclude

$$\Phi \geq 1 - 4q - 2q = 1 - 6q \quad (13)$$

If the extension of land is sufficiently small, $1 - 6q > 0$, and the proof is completed.

3. The Modified Altimetry-Gravimetry Problem 1

The calibration of the altimeter can be done nowadays to a very satisfactory level of accuracy, *i.e.*, $\pm(2-3)\text{cm}$. We also believe that one cannot assume known anywhere on earth the absolute gravitational potential, as only potential differences are accurately observable. Nevertheless, we observe that the model (4) is derived by linearization under the following assumptions:

- on land, potential differences $W(P) - W_0$ are known, where, for instance, W_0 could be considered as the actual gravity potential at a reference tide gauge point;
- on land, furthermore, the modulus g of the gravity vector is known everywhere;

- on sea, from a good circulation model, the height anomaly $\zeta(P)$ with respect to some equipotential reference surface is known.

As a matter of fact, the most simple and universally accepted stationary model for the ocean is what is called the geostrophic balance model, claiming that

$$\nabla \zeta = \frac{2\Omega \sin \phi}{g} \mathbf{n} \times \mathbf{v} \quad (14)$$

where \mathbf{n} is the local vertical, \mathbf{v} is the horizontal velocity of the geostrophic current, Ω is the earth's rotation angular rate, and ϕ is the latitude. A circulation model can provide \mathbf{v} , therefore equation (14) can be integrated and ζ achieved but for a constant.

Model (14) is actually known to give a satisfactory approximation of mid-ocean circulation (certainly it does not hold on the western boundary — cf. Pedlosky, 1996; Wunsch, 1993), and, more generally, no universally accepted oceanographic model is available to describe in detail the effects of stationary currents along irregular coasts. It follows that indeed in the part of the ocean described by (14) it can be asserted that, for any couple of points

$$W_1 + \gamma_1 \zeta_1 = W_2 + \gamma_2 \zeta_2 = \bar{W} \quad (15)$$

and consequently

$$W_1 - W_2 = \gamma_2 \zeta_2 - \gamma_1 \zeta_1 \simeq \bar{\gamma}(\zeta_2 - \zeta_1) \quad (16)$$

The potential is known, in other words, but for an unknown additive constant. If such a model could be extended to the tide gauge where the potential reference point for land is located, it could be asserted that *one and the same* W_0 could be the only unknown constant. The linearization of such a problem is provided by equation (4). Nevertheless, if such a connection is not established, the problem should rather be formulated as

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \\ T = f_1 + \delta W_1 & \text{on } S_S \\ -\frac{\partial T}{\partial r} - \frac{2}{r} T = f_2 - \frac{2}{r} \delta W_2 & \text{on } S_L \\ \int_{S_L} T d\sigma = 0 & \end{cases} \quad (17)$$

Having introduced two unknown constants, however, one should be working with two constraints. So, in addition to $\int_{S_L} T d\sigma = 0$, expressing that normal and actual potential are generated by the same mass quantity M , another constraint has to be introduced, which can be chosen to be

$$\int_{S_S} T d\sigma = 0 \quad (18)$$

meaning that the mean equipotential surface interpolating the actual surface of the ocean (given by altimetry) is chosen as reference in the linearization process

(Sacerdote and Sansò, 2001). In other words, it is acceptable to have different height datums on the oceans and on the continents, provided that the boundary-value problem theory is capable of establishing a well defined relation between them.

The proof of the well-posedness of problem (17), (18) for a spherical boundary of radius R is very simple; it relies again on the Fredholm alternative and on the uniqueness of the solution of the corresponding homogeneous problem:

$$\begin{cases} \nabla^2 T = 0 & \text{in } \Omega \\ T = \delta W_1 & \text{on } S_S \\ -\frac{\partial T}{\partial r} - \frac{2}{R} T = -\frac{2}{R} \delta W_2 & \text{on } S_L \\ \int_S T d\sigma = 0, \int_{S_S} T d\sigma = 0 \end{cases} \quad (19)$$

First of all, from the constraints in equation (19) it follows immediately that $\int_{S_L} T d\sigma = 0$ too. Integrating the boundary values over the spherical boundary surface, and using spherical harmonic expansions yields

$$\frac{1}{4\pi} \int_S -\left(\frac{\partial T}{\partial r} + \frac{2}{R} T\right) T d\sigma = \sum \frac{n-1}{R} T_{nm}^2 \quad (20)$$

On the other hand, introducing boundary conditions, and recalling that, by virtue of the conditions in equation (19), both $\int_{S_L} T d\sigma$ and $\int_{S_S} T d\sigma$ vanish, one obtains

$$\begin{aligned} \frac{1}{4\pi} \int_S -\left(\frac{\partial T}{\partial r} + \frac{2}{R} T\right) T d\sigma &= -\frac{\delta W_1}{4\pi} \int_{S_S} \left(\frac{\partial T}{\partial r} + \frac{2}{R} T\right) d\sigma \\ &\quad - \frac{2}{R} \frac{\delta W_2}{4\pi} \int_{S_L} T d\sigma \\ &= -\frac{\delta W_1}{4\pi} \int_{S_S} \frac{\partial T}{\partial r} d\sigma \end{aligned} \quad (21)$$

From equation (19) it can be immediately inferred that

$$0 = \int_{S_S} T d\sigma = \delta W_1 \mu(S_S) \quad (22)$$

from which $\delta W_1 = 0$. Consequently, the last term in equation (21) vanishes, and from equations (20) and (21) it follows that $\sum \frac{n-1}{R} T_{nm}^2 = 0$. Furthermore from (19) $T_{00} = \int_S T d\sigma = 0$. Hence $\sum \frac{n-1}{R} T_{nm}^2$ must be strictly positive, unless all T_{nm} with $n \neq 1$ vanish. On the other hand, no degree-1 harmonic function can vanish identically on a portion of a spherical surface, as required by the boundary condition on S_S in equation (19), together with equation (22); consequently the coefficients T_{1m} must vanish as well. Finally, also δW_2 must vanish as a consequence of the third of the equations (19).

This procedure can be generalized to a more realistic situation, when the land portion is divided into a number of separate continents, which cannot be altimetrically connected. In this case a larger number of different unknown constants must be introduced, which must be balanced by a corresponding number of constraint equations, in order to make the problem uniquely solvable.

4. Concluding Remark

If physical knowledge based on measurements and oceanographic models is properly taken into account, a modified altimetry-gravimetry problem may be formulated, the mathematical properties of which look very “pleasant” and classical. This result is quite satisfactory, and is a good representation of the actual capability of geodesy to retrieve the disturbing gravity potential from altimetry-gravimetry boundary data.

References

- Heiskanen, W. and Moritz, H.: 1967, *Physical Geodesy*, W.H. Freeman & Co., San Francisco, 364 pp.
- Lehmann, R.: 2000, ‘Altimetry-Gravimetry Problems with Free Vertical Datum’, *Journal of Geodesy* **74**, 327–334.
- Pedlosky, J.: 1996, *Ocean Circulation Theory*, Springer Verlag, Heidelberg, 453 pp.
- Sacerdote, F. and Sansò, F.: 1987, ‘Further Remarks on the Altimetry-Gravimetry Problems’, *Bull. Géod.* **61**, 65–82.
- Sacerdote, F. and Sansò, F.: 2001, ‘ W_0 : A story of the height datum problem’, in *Festschrift W. Torge, Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover* **241**, 49–56.
- Svensson, L.: 1985, ‘Some Remarks on the Altimetry-Gravimetry Problem’, *Proc. of the I Hotine-Marussi Symposium on Math. Geodesy*, 559–582.
- Svensson, L.: 1991, ‘The Boundary Equation Approach to the Geoid’, in (R. H. Rapp and F. Sansò (eds.), *Determination of the Geoid - Present and Future*, IAG Symposia 106, Springer-Verlag, pp. 357–364.
- Wunsch, C.: 1993, ‘Physics of the Ocean Circulation’, in R. Rummel and F. Sansò (eds.), *Satellite Altimetry in Geodesy and Oceanography, Lecture Notes in Earth Sciences* **50**, Springer-Verlag, pp. 9–98.

OCEAN TIDES IN GRACE MONTHLY AVERAGED GRAVITY FIELDS

PER KNUDSEN

Kort & Matrikelstyrelsen, Geodetic Department, Denmark (pk@kms.dk)

Received: 28 August 2002; Accepted in final form: 10 October 2002

Abstract. The GRACE mission will map the Earth's gravity fields and its variations with unprecedented accuracy during its 5-year lifetime. Unless ocean tide signals and their load upon the solid earth are removed from the GRACE data, their long period aliases obscure more subtle climate signals which GRACE aims at. In this analysis the results of Knudsen and Andersen (2002) have been verified using actual post-launch orbit parameter of the GRACE mission. The current ocean tide models are not accurate enough to correct GRACE data at harmonic degrees lower than 47. The accumulated tidal errors may affect the GRACE data up to harmonic degree 60. A study of the revised alias frequencies confirm that the ocean tide errors will not cancel in the GRACE monthly averaged temporal gravity fields. The S_2 and the K_2 terms have alias frequencies much longer than 30 days, so they remain almost unreduced in the monthly averages. Those results have been verified using a simulated 30 days GRACE orbit. The results show that the magnitudes of the monthly averaged values are slightly higher than the previous values. This may be caused by insufficient sampling to fully resolve and reduce the tidal signals at short wavelengths and close to the poles.

Keywords: GRACE satellite mission, ocean tides, gravity

1. Introduction

The GRACE mission will accurately map variations in the Earth's gravity field over its 5-year lifetime (Tapley, 1997; Bettadpur *et al.*, 1999). It will provide scientists from all over the world with an efficient way to map the Earth's gravity fields with unprecedented accuracy. The results from this mission will yield crucial information about the distribution and flow of mass within the Earth and it's surroundings. The gravity variations that GRACE will study include: changes due to surface and deep currents in the ocean; runoff and ground water storage on land masses; exchanges between ice sheets or glaciers and the oceans; and variations of mass within the Earth (Wahr *et al.*, 1998). To reach these goals it is important to eliminate the effects of ocean tides on the estimated gravity field. Hence, an evaluation of the quality of existing models is important.

Previous studies have shown that ocean tides are important to consider and that the ocean tide models are not accurate enough to correct GRACE gravity fields fully at lower harmonic degrees (Knudsen *et al.*, 2000; Ray *et al.*, 2001). Furthermore, the results of the previous analysis indicate that ocean tides, e.g. the S_2 term, will not cancel out in the monthly averaged gravity fields due to the sampling characteristics of GRACE (Knudsen *et al.*, 2000).



Space Science Reviews **108:** 261–270, 2003.

© 2003 Kluwer Academic Publishers. Printed in the Netherlands.

In Knudsen and Andersen (2002) procedures for correction of GRACE gravity fields for effects by ocean tides are revised. The accuracies of the M_2 , S_2 , K_1 , and O_1 ocean tide constituents of the recent and highly accurate models NAO99 (Matsumoto *et al.*, 2000) and CSR4.0 (Eanes and Bettadpur, 1995) are assessed using spherical harmonic functions and the expected error degree variances for GRACE (Bettadpur, personal communication). The sampling characteristics are revised and preliminary alias frequencies associated with eight tidal constituents are computed to analyse how tidal errors affect the GRACE monthly temporal averaged gravity fields.

The results of Knudsen and Andersen (2002) confirm that the current ocean tide models are not accurate enough to correct GRACE data at harmonic degrees lower than 35. The accumulated tidal errors may affect the GRACE data up to harmonic degree 56. The study of the approximate alias frequencies show that the ocean tide errors will not cancel in the GRACE monthly averaged temporal gravity fields. The S_2 and the K_2 terms have alias frequencies much longer than 30 days, so they remain almost unreduced in the monthly averages.

In this analysis the effects of ocean tides in the monthly averaged GRACE gravity fields are further studied by using the newly available actual orbit parameter and a simulated 30 days GRACE orbit. Furthermore, the results in Knudsen and Andersen (2002) are revised using new harmonic expansions of the tidal constituents, since the original harmonic expansions were found to be affected by a smoothing regularization algorithm.

2. Gravity Changes due to Ocean Tides

The gravity changes due to ocean tides are caused by the changes in mass that follow the ocean tides. At the ocean surface the gravity change can be expressed as a free air anomaly change using the Bouguer plate approximation as (*i.e.*, Telford *et al.*, 1976)

$$\Delta g = 0.0434H(t) \quad (1)$$

where $H(t)$ is the height of the sea surface above the ocean bottom varying due to ocean tides and the anomaly change is in units of mGal (10^{-5} m/s 2). In (1) the height is in meters and a standard value for the ocean water density of 1.036 g/cm 3 is used.

The ocean tides consist of several constituents associated with different combinations of the attractions from the sun and the moon. That is

$$H(t) = \sum_k H_k = \sum_k U_k \cos(\omega_k t + \chi_k) + V_k \sin(\omega_k t + \chi_k) \quad (2)$$

where U_k and V_k are coefficients, ω_k is the tidal frequency, and χ_k is the astronomical argument associated with the k 'th constituent. From the coefficients the amplitude A_k can be obtained using

$$A_k^2 = U_k^2 + V_k^2. \quad (3)$$

The RMS values of the respective variations are obtained from the amplitudes by dividing by the square root of two.

Parts of this analysis are carried out using spherical harmonic functions. Hence, each of the U_k and V_k quantities are expanded in spherical harmonic functions. That is for an arbitrary surface S

$$S(\phi, \lambda) = \sum_{l=1}^L \sum_{m=0}^l [C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda)] P_{lm}(\cos \phi) \quad (4)$$

where P_{lm} is the fully normalised associated Legendre function of degree and order l and m respectively. Subsequently the degree variances are

$$\sigma_l = \sum_{m=0}^l [C_{lm}^2 + S_{lm}^2]. \quad (5)$$

For each tidal constituent the degree variances associated with the gravity anomalies are obtained using the spherical harmonic expansion, (4) of both the U and the V surfaces of that tidal constituent, (2), converted to gravity using (1) and, finally, combined using (5).

3. Ocean Tide Correction Accuracy

The accuracy of GRACE can be evaluated at different wavelengths by using the expected error degree variances for GRACE (Bettadpur, personal communication). Converted into free air anomalies those degree variances may be compared directly with changes in height through (1). In Knudsen and Andersen (2002) such errors or accuracy estimates associated with different harmonic degrees have been described. At harmonic degree 20 the accumulated GRACE error is 0.043 micro Gal corresponding to a sea level change of one millimetre. At harmonic degree 100 the accumulated GRACE error is about one thousand times larger. Those and other estimates listed in Table I (from Knudsen and Andersen, 2002) demonstrate the different requirements to the accuracy of the correction, *e.g.* at harmonic degree 50 an accuracy of 1.2 centimetre is required.

To evaluate the errors in the tide models Knudsen and Andersen (2002) used the differences between the hydrodynamic model NAO99 (Matsumoto *et al.*, 2000) and the latest altimetry enhanced model CSR-4.0 (Eanes and Bettadpur, 1995) for the M_2 , S_2 , K_1 , and the O_1 constituents. Both of these models are based on

TABLE I

Expected accumulated error of GRACE expressed as free air anomalies in micro Gal together with corresponding change in sea level height in centimetres (from Knudsen and Andersen, 2002).

Harmonic Degree	Accumulated Gravity Error	Change in Sea-level Height
20	0.04	0.1
30	0.10	0.2
40	0.24	0.6
50	0.53	1.2
60	1.29	3.0
80	7.34	16.9
100	51.96	119.7

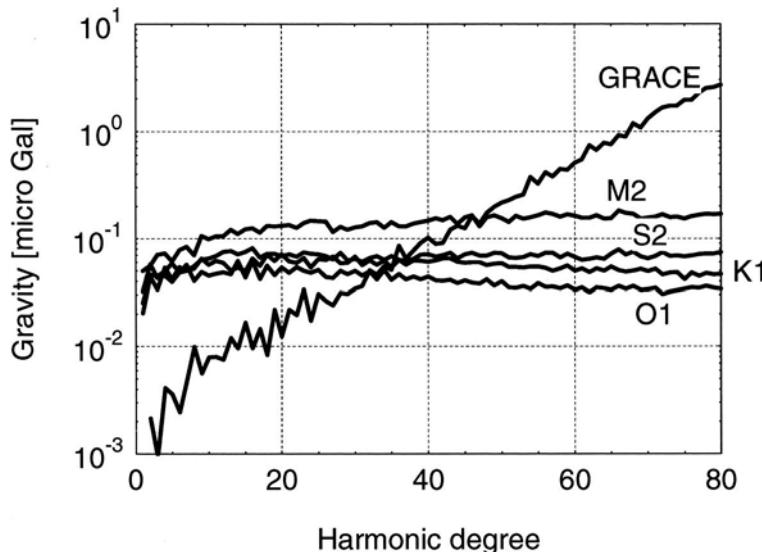


Figure 1. Error degree variances of the M_2 , S_2 , K_1 , and O_1 tidal models compared with the accuracies of GRACE (revised version of Figure 1 in Knudsen and Andersen, 2002).

TOPEX/POSEIDON altimetry, so the difference may underestimate the total tidal error (Ray *et al.*, 2001). The results by Knudsen and Andersen (2002) were found to be affected by smoothing due to a regularization that was applied in the computation of the spherical harmonic expansions. For this reason new harmonic expansions of the differences were computed and the degree variances associated with free air gravity anomalies were derived. The new results (Figure 1) show that the M_2 , S_2 , K_1 , and the O_1 constituents are significant up to harmonic degrees 47, 35,

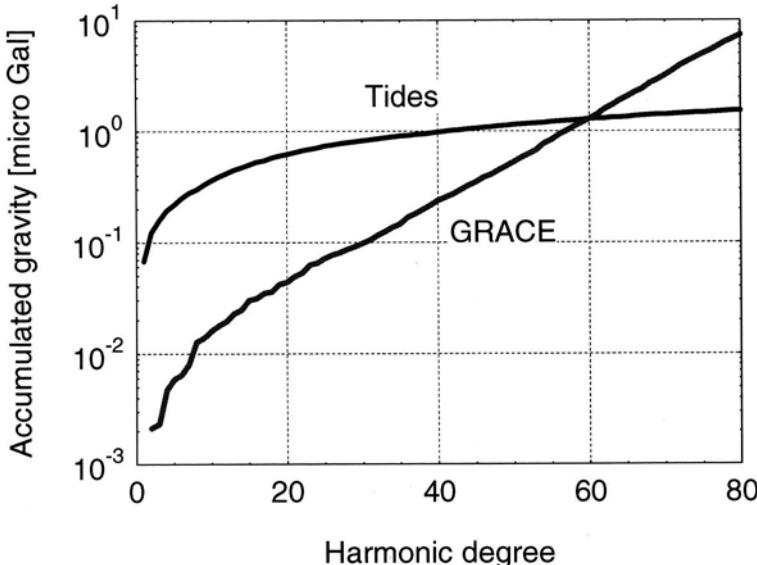


Figure 2. Accumulated error degree variances of the combined M_2 , S_2 , K_1 , and O_1 tidal models compared with the accumulated error of GRACE. At harmonic degree 60 the accumulated curves have a value of 1.3 micro Gal corresponding to a sea level change of 3.0 cm (revised version of Figure 2 in Knudsen and Andersen, 2002).

35, and 32 respectively. The sum of the tidal errors from all four constituents accumulated by harmonic degree is shown in Figure 2 together with the accumulated error of GRACE. As in Knudsen and Andersen (2002) the accumulated error of the tidal constituents has reached a value of about one micro Gal Above harmonic degree 40. Now it intersects the increasing accumulated GRACE error at harmonic degree 60. Hence, the sum of the tidal errors may be significant in GRACE data up to harmonic degree and order 60.

4. Alias Frequencies of the Tidal Constituents

The errors described in this study will affect the single measurements of GRACE. How the errors will affect GRACE products (*e.g.* models of monthly averaged temporal gravity fields) depend on the sampling frequency of GRACE, which depends on the orbit configuration for GRACE. As pointed out by Knudsen and Andersen (2002) the GRACE orbit will be very stable in the inertial system due to its high inclination, $i = 89^\circ$ (*e.g.*, Kaula, 1966). Hence, the solar tides will appear as a signal with an alias frequency close to annual or semiannual cycles. Subsequently, the errors in the models of the solar tides will not be averaged out over 30 day intervals. By using the actual orbit parameters the actual precession of the node,

TABLE II

Tidal frequencies and preliminary alias frequencies assuming a sampling frequency of 0.49846 days in days. Furthermore, modulations caused by the sampling of diurnal constituents are shown in days (revised version of Table 2 in Knudsen and Andersen, 2002).

Constituent	Frequency	Alias	Modulation
M ₂	0.5175	13.6	—
S ₂	0.5000	162.2	—
N ₂	0.5274	9.1	—
K ₂	0.4986	1460	—
K ₁	0.9973	0.9969	2920
O ₁	1.0758	0.9969	13.6
P ₁	1.0028	0.9969	9.1
Q ₁	1.1195	0.9969	9.1

omega dot, was computed and is -0.12 degrees per day, which is about -45 degrees per year only.

To study the characteristics of the ocean tides as sampled by GRACE in detail, alias frequencies of the eight largest tidal constituents were computed (*e.g.*, Knudsen, 1994; Knudsen and Andersen, 2002). As in Knudsen and Andersen (2002) a sampling interval close to half a sidereal day was assumed. This corresponds to a sampling of the gravity field at both ascending and descending tracks, which will be relevant except for areas near the poles. In the analysis the actual precession of the node was taken into account. Hence, a sampling interval of 0.49846 days was applied in this analysis. The GRACE satellite will fly in a non-repeating orbit that complicates the definition of alias frequencies, since the sampling will not be regular (Knudsen and Andersen, 2002). However, GRACE will measure the gravity field averaged over an area of a few hundred kilometres. Considering such a region the satellite may sample the gravity field several times during a one-month period at times separated by multiples of the assumed sampling frequency.

The results of the computation of alias frequencies (see Table II) are slightly different from the results by Knudsen and Andersen (2002). They show that M₂ has an alias frequency of 13.6 days and that S₂ has an alias frequency of 162.2 days. N₂ has an alias frequency of 9.1 days. K₂ has a frequency closer to the sampling, so K₂ has an alias frequency of 1460 days or four years. The diurnal constituents are sampled sufficiently to avoid aliases. However, their frequencies are not identical to the sampling, so a modulation of the amplitude will appear. For K₁, O₁, P₁, and Q₁

TABLE III

Relative magnitudes of tidal signal averaged over one month, both a theoretical sinc-values and actual values from simulations (revised version of Table 2 in Knudsen and Andersen, 2002).

Constituent	Alias Frequency	Averaged	
		sinc	actual
M ₂	13.6	0.10	0.07
S ₂	162.2	0.94	0.95
N ₂	9.1	0.08	–
K ₂	1460	1.00	–
K ₁	0.9969	0.01	0.07
O ₁	0.9969	0.01	0.04
P ₁	0.9969	0.01	–
Q ₁	0.9969	0.01	–

the amplitudes will be modulated by frequencies of 2920, 13.6, 9.1, and 9.1 days respectively.

5. Tidal Errors in GRACE Monthly Averages

In Knudsen and Andersen (2002) the effect of averaging GRACE gravity over monthly intervals on the tidal errors was analysed using the alias frequencies. In the frequency domain a convolution by a block averaging function corresponds to a multiplication by a sinc function ($\text{sinc}(u) = \sin(u)/u$). Using this function together with the revised alias frequencies the magnitudes of the monthly averaged tidal constituents were recomputed (see Table III). The effect of averaging GRACE gravity over monthly intervals on the tidal errors show the results that M₂, S₂, N₂, and K₂ are reduced to 10%, 94%, 8%, and 100% of their original magnitudes respectively. Hence, the results by Knudsen and Andersen (2002) have been confirmed using the actual orbit parameters. S₂ and K₂ are practically unreduced in the GRACE monthly averaged gravity fields. The diurnal tides are almost fully reduced.

As pointed out by Knudsen and Andersen (2002), GRACE will fly in a non-repeating orbit that complicates the definition of alias frequencies as well as the magnitude of the monthly averages. To verify the results a simulated 30 days GRACE orbit was constructed using a circular orbit with a height of 500 km and an inclination of 89°. The period of one revolution is 5679 seconds. Hence, in

one day the satellite accomplish about 15.21 revolutions. Then using the positions and times of points separated by 2 degrees along the orbit the actual tidal phases represented by both the cosine and sine terms (2), were computed for the M_2 , S_2 , K_1 , and the O_1 constituents.

An initial inspection of the synthetic data was carried out using the phases at equator crossings of the 30 days simulated orbit. Those are shown in Figure 3–4. The figures show the sampling pattern along the equator with time. In Figure 3 the M_2 phases and the S_2 phases are shown. For both constituents the phases are quite consistent regionally. Through the 30 days change in time the M_2 phases change by more than two cycles. In contrast, the S_2 phases does not change much through the 30 days period. Those results indicate that the previous analysis of the alias frequencies is relevant. In Figure 4 the K_1 phases and the O_1 phases are shown. In contrast to the semidiurnal tides shown in Figure 3, the phases of the diurnal tides are not very consistent regionally. Between ascending and descending tracks the phases change by about 180 degrees. This result also agrees with the previous results.

Then spherical harmonic functions complete to harmonic degree and order 80 were fitted to cosine and the sine terms respectively, for each of the four tidal constituents. Then, for each tidal constituent, surfaces associated with the cosine and the sine terms, or the U and V quantities of (2), covering the earth, were computed. Subsequently, amplitudes were derived using (3) and averaged over the surface of the earth. The results shown (Table III) that M_2 and S_2 are reduced to 7% and 95% respectively, of their original magnitudes. Those results agree very well with the previous results. The K_1 and the O_1 terms are reduced to 7% and 4% respectively. Those values are slightly higher than the previous values, which may be caused by insufficient sampling to fully resolve and reduce the tidal signals at short wavelengths and close to the poles.

6. Discussion

The results of Knudsen and Andersen (2002) have been confirmed using actual orbit parameter of the GRACE mission. The current ocean tide models are not accurate enough to correct GRACE data at harmonic degrees lower than 47. The accumulated tidal errors may affect the GRACE data up to harmonic degree 60. A study of the revised alias frequencies show that the ocean tide errors will not cancel in the GRACE monthly averaged temporal gravity fields. The S_2 and the K_2 terms have alias frequencies much longer than 30 days, so they remain almost unreduced in the monthly averages.

Those results have been verified using a simulated 30 days GRACE orbit, where positions and times of points along the orbit were used together with the cosine and sine terms for the M_2 , S_2 , K_1 , and the O_1 constituents. Then spherical harmonic functions complete to harmonic degree and order 80 were fitted to cosine and the

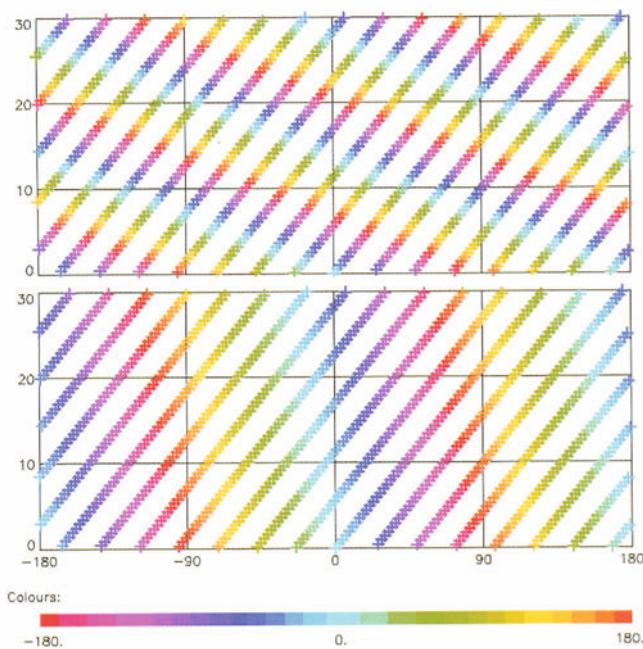


Figure 3. M₂ phases (upper) and S₂ phases (lower) at longitudes and times of equator crossings of a 30 days simulated orbit.

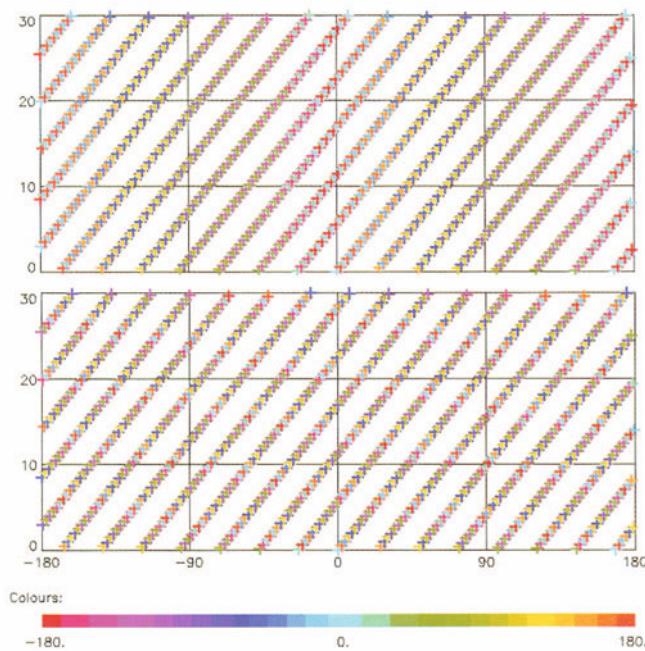


Figure 4. K₁ phases (upper) and O₁ phases (lower) at longitudes and times of equator crossings of a 30 days simulated orbit.

sine terms respectively, for each of the four tidal constituents. The results show that the magnitudes of the monthly averaged values are slightly higher than the previous values. This may be caused by insufficient sampling to fully resolve and reduce the tidal signals at short wavelengths and close to the poles.

At harmonic degrees lower than 20 sea level signals as low as one millimetre will be significant in the GRACE temporal gravity fields. Hence, errors in both the ocean tide models and the loading models need to be considered. Spatial and temporal changes in the ocean water density may also be significant. On the other hand, the GRACE mission provide a new opportunity to improve the ocean tide models, which in turn will improve the modelling of the distribution and flow of mass within the Earth and its surroundings.

Acknowledgements

This analysis is a contribution to the GRACE project as a result of studies within the GRACE science team.

References

- Bettadpur, S., Gruber, Th., and Watkins, M.: 1999, 'GRACE Science Data System Design', *Supplement to EOS Transactions, AGU Fall Meeting* **80**(46), F254.
- Eanes, R., and Bettadpur, S.: 1995, 'The CSR 3.0 global ocean tide model, CSR-TM- 95-06', *Technical Memorandum*, Center for Space Research, December 1995.
- Kaula, W. M.: 1966, 'Theory of Satellite Orbits', Blaisdell Publishing Co., Waltham, Mass, 124 pp.
- Knudsen, P.: 1994, 'Separation of residual ocean tide signals in a collinear analysis of GEOSAT altimetry', *Bull. Geod.* **68**, 7–18.
- Knudsen, P., and Andersen, O.: 'Correcting GRACE gravity fields for ocean tide effects', *Geophys. Res. Lett.* **29**(8), 19-1–19-4.
- Knudsen, P., Andersen, O. B., Khan, S. A., and Høyer, J. L.: 2000, 'Ocean Tide Effects on GRACE Gravimetry', *IAG Symposia, Gravity Geoid and Geodynamics 2000*, Vol. **123**, Springer-Verlag, pp. 159–164.
- Matsumoto, K., Takanezawa, T., and Ooe, M.: 2000, 'Ocean Tide Models Developed by Assimilating TOPEX/POSEIDON Altimeter Data into Hydrodynamical Model: A Global Model and a Regional Model Around Japan', *Journal of Oceanography* **56**, 567–581.
- Ray, R. D., Eanes, R. J., Egbert, G. D., and Pavlis, N. K.: 2001, 'Error spectrum for the global M_2 ocean tide', *Geophys. Res. Lett.* **28**, 21–24.
- Tapley, B. D.: 1997, 'The gravity recovery and climate experiment (GRACE)', *Supplement to EOS Trans. AGU* **78**(46), pp. 163.
- Telford, W. M., Geldart, L. P., Sheriff, R. E., Keys, D. A.: 1976, *Applied Geophysics*, Cambridge University Press.
- Wahr, J., Molenaar, M., and Bryan, F.: 1998, 'Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE', *J. Geophys. Res.* **103**(B12), 30205–30229.

TIDAL MODELS IN A NEW ERA OF SATELLITE GRAVIMETRY

R. D. RAY and D. D. ROWLANDS

Space Geodesy Branch, NASA Goddard Space Flight Center, Greenbelt, USA

G. D. EGBERT

College of Oceanic & Atmospheric Sciences, Oregon State University, Corvallis, USA

Received: 15 May 2002; Accepted in final form: 20 December 2002

Abstract. The high precision gravity measurements to be made by recently launched (and recently approved) satellites place new demands on models of Earth, atmospheric, and oceanic tides. The latter is the most problematic. The ocean tides induce variations in the Earth's geoid by amounts that far exceed the new satellite sensitivities, and tidal models must be used to correct for this. Two methods are used here to determine the standard errors in current ocean tide models. At long wavelengths these errors exceed the sensitivity of the GRACE mission. Tidal errors will not prevent the new satellite missions from improving our knowledge of the geopotential by orders of magnitude, but the errors may well contaminate GRACE estimates of temporal variations in gravity. Solar tides are especially problematic because of their long alias periods. The satellite data may be used to improve tidal models once a sufficiently long time series is obtained. Improvements in the long-wavelength components of lunar tides are especially promising.

1. Introduction

With the recent launches of the CHAMP and GRACE satellite missions and the approval of the GOCE mission, we find ourselves entering a new era in satellite gravimetry. The high precisions and increased sensitivities of these missions place new demands on methodologies and models. This paper examines the role of tides and tide model errors, with special emphasis given to the GRACE low-low satellite-to-satellite tracking mission (Tapley and Reigber, 2000).

Tides have been a key component of space geodesy since the mid-1960's, but their role in satellite gravimetry has been confined primarily (although not exclusively) to determining long-period tidal perturbations in satellite orbits or to estimating certain tidal parameters from such perturbations. Determining long-period orbit perturbations will continue to be of utmost importance, in these new missions as well as in other high-precision applications (e.g., Pavlis and Iorio, 2002). But in this new era, attention will also turn to two new aspects (cf. Schrama, 1996). (1) The continuous GPS tracking of satellites now allows short-period tidally induced orbit perturbations to be studied, necessitating models with complete sets of spherical harmonic coefficients, not limited to a few select orders. (2) Some of the new data types allow a direct instantaneous measurement of the gravity effects of tides (embedded, of course, in a background of gravity signals from other



anomaly sources). Unlike long-period orbit perturbations, these direct* effects can be localized near the causative masses. Furthermore, the contamination of gravity field measurements (or estimates) from inadequately modeled tidal masses will be an important concern. In this paper we consider the gravitational effects of tides and examine how errors in tidal models may affect the new satellite gravimetry missions. Space precludes a complete discussion, so we focus on oceanic tides, as opposed to Earth tides or atmospheric tides, because errors in the former are much larger than errors in the latter.

2. Ocean Tides and Gravity

To understand the extent to which new gravity missions are directly affected by ocean tides, we use a global model to evaluate (1) the tidally induced perturbations in gravity (or the geoid) and (2) the associated signals that can be expected in satellite measurements. Later sections consider corresponding errors in present-day models.

2.1. GRAVITY PERTURBATIONS INDUCED BY OCEAN TIDES

Suppose the ocean tidal elevations are expanded in normalized complex spherical harmonics $Y_n^m(\theta, \varphi)$ as

$$\zeta(\theta, \varphi, t) = \sum_{n,m} z_{nm}(t) Y_n^m(\theta, \varphi). \quad (1)$$

The coefficients $z_{nm}(t)$ vary with tidal periodicity. At satellite altitudes the gravitational potential of this tide is given by (e.g., Lambeck, 1988)

$$U(r, \theta, \varphi, t) = 4\pi G a \rho_w \sum_{n,m} \frac{1 + k'_n}{2n + 1} \left(\frac{a}{r}\right)^{n+1} z_{nm}(t) Y_n^m(\theta, \varphi) \quad (2)$$

where ρ_w is the mean density of seawater (approximately 1035 kg m^{-3}), a the radius of the Earth, G is the Newtonian constant, and k'_n are loading Love numbers.

The corresponding induced variations in the geoid are given by (e.g., Schrama, 1996; Wahr, 1998)

$$\delta N(\theta, \varphi, t) = 3(\rho_w/\rho_e) \sum_{n,m} \left(\frac{1 + k'_n}{2n + 1}\right) z_{nm}(t) Y_n^m(\theta, \varphi), \quad (3)$$

where ρ_e is the mean density of the Earth. Figure 1 shows the amplitudes of both ζ and δN for the principal semidiurnal lunar tide M_2 .

* Here ‘direct’ is used in contrast with the indirect tidal gravity effect manifested by orbit perturbations. Schrama (1996) prefers ‘indirect’ for both and reserves ‘direct’ for the primary astronomical tidal forces of the sun or moon. The latter are exceedingly well known and not discussed in this paper.

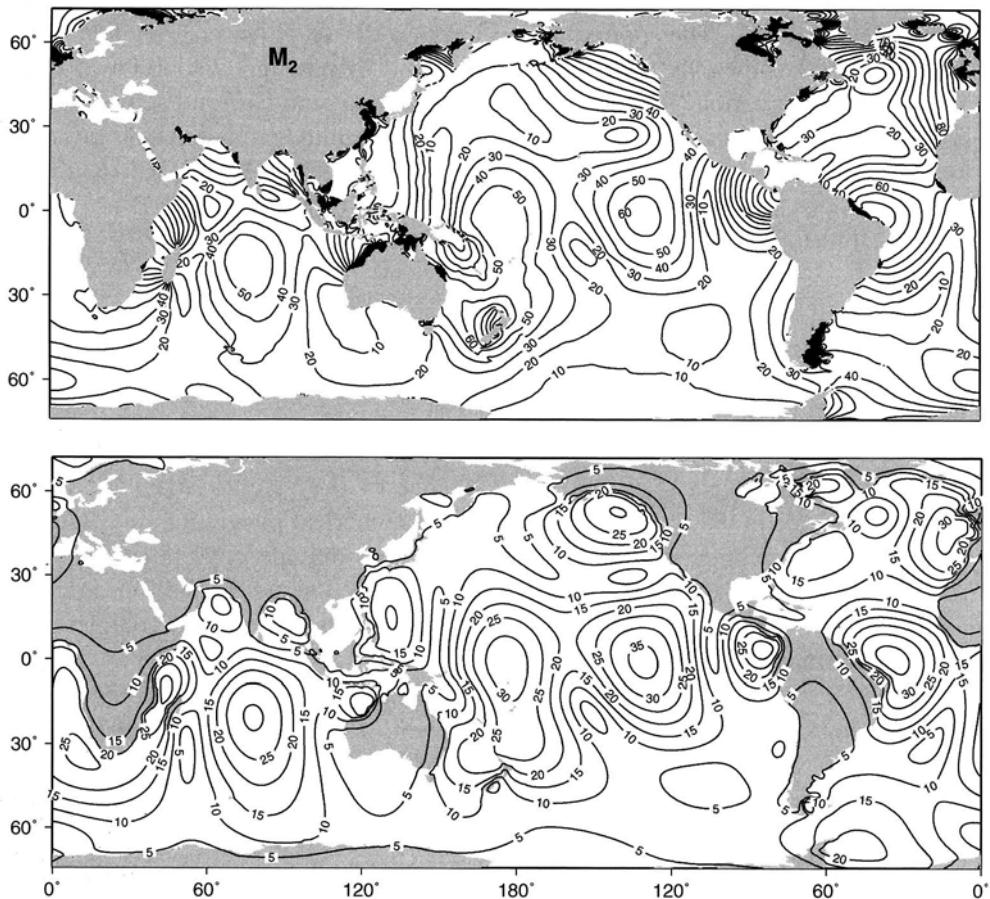


Figure 1. Amplitude of the M_2 ocean tide (top) in cm and the corresponding induced geoid perturbation (bottom) in mm. The ocean model is based primarily on Topex/Poseidon data. The geoid perturbation follows from Eq. (3).

For reference the tidal perturbations in the dimensionless Stokes coefficients C_{nm} are related to the elevation coefficients by

$$\delta C_{nm}(t) = \frac{3\rho_w}{a\rho_e} \frac{(1+k'_n)}{(2n+1)} z_{nm}(t). \quad (4)$$

At radius $r = a$ the tidal perturbations in the gravity field $-\partial U/\partial r$ are

$$\delta g(\theta, \varphi, t) = 4\pi G\rho_w \sum_{n,m} (1+k'_n) \left(\frac{n+1}{2n+1} \right) z_{nm}(t) Y_n^m(\theta, \varphi). \quad (5)$$

For sufficiently large n the gravity perturbation is related to the tidal height simply by

$$\lim_{n \rightarrow \infty} \delta g_{nm} = 2\pi G\rho_w z_{nm} \approx (0.0434 \text{ mGal/m}) z_{nm},$$

which leads to the Bouguer plate approximation ($\delta g \approx 0.0434\zeta$) used by Knudsen and Andersen (2002). The approximation works well except for very small n .

As Figure 1 shows, the tides perturb the geoid by amounts well below 5 cm. The largest perturbation (for M_2) is 3.7 cm in the equatorial eastern Pacific. Geoid signals of this magnitude are about an order of magnitude smaller than present uncertainties in the geoid over the ocean (Lemoine et al., 1998), but they are well within the anticipated precisions of GRACE and GOCE (for GOCE, see ESA [1999] Table 8.4).

2.2. OCEAN TIDE SIGNALS IN GRACE AND GOCE

The gravitational signals of ocean tides relevant to the GRACE and GOCE measurements can be predicted with a global ocean tide model. Cheng (2002) recently employed linear perturbation methods to compute the anticipated spectrum of tidal signals in GRACE data over a frequency range of 0–150 cycles/day. We here show an example in the time domain based on a numerical integration of GRACE ephemerides (including initial state adjustments) at altitude approximately 500 km. The relevant forcing from ocean tides is computed using the four largest constituents of the GOT00.2 ocean model (an update to Ray [1999]) complete to spherical harmonic degree 60.

Figure 2a shows the expected ocean tide signal in GRACE’s range-rate measurement for an arbitrary 120-minute arc segment, or about 1.5 revolutions. Most of this signal should be observable with GRACE’s 1- $\mu\text{m/s}$ precision, especially given that these signals are temporally coherent.

As an alternative to employing numerical integration or linear perturbation methods, it is sometimes useful to approximate these range rate signals by employing the following simplistic method. For a perfectly circular and polar orbit, the along-track acceleration is given by $\hat{\theta} \cdot \nabla U$ and the velocity thus by $(1/r) \int (\partial U / \partial \theta) dt$. Within the integrand we approximate $\dot{\theta}$ as simply $(GM/r^3)^{1/2}$, with M the Earth’s mass, so that the along-track velocity integrates to $v(t) \approx (r/GM)^{1/2}U(t)$. The range rate \dot{r} is then taken as the difference $v(t) - v(t + \Delta t)$, with Δt the time lag between the two satellites, or

$$\dot{r}(t) \approx (r/GM)^{1/2}(U(t) - U(t + \Delta t)). \quad (6)$$

Of course, this simple approximation fails to account for the dynamically resonant terms in range rate at 0, 1, and 2 cycles/rev; these terms, which in fact are dominant for GRACE (Cheng, 2002), can only be recovered by integration of the equations of motion. Nonetheless, the method usefully captures the shorter period signals in range-rate, as can be seen in Figure 2b. The difference between Figure 2a and 2b is predominantly a sum of two sinusoids of frequency 1 and 2 cycles/rev.

The planned GOCE satellite will measure directly the nine components of the second-order gravity gradient tensor. Figure 3 shows the predicted tidal signal in the second-order radial gradient, which is generally the largest component (Schrama,

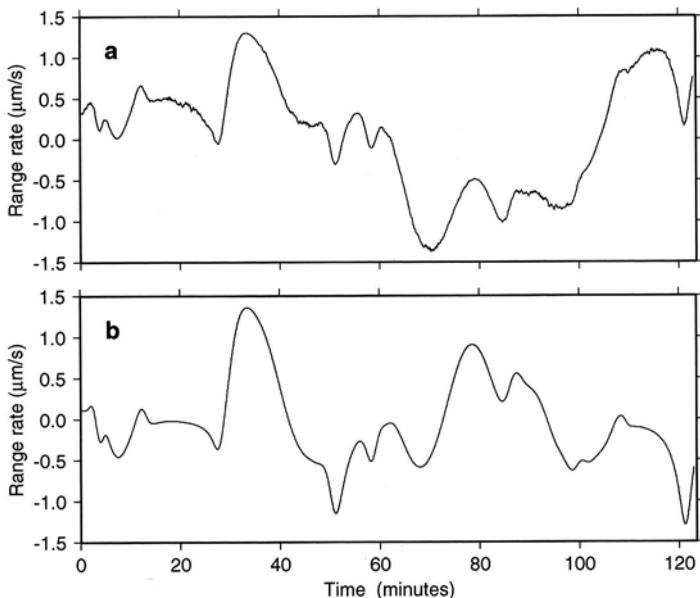


Figure 2. Predicted ocean tide signals in the GRACE range rate measurements for an arbitrary 120 minutes of a 1-day arc, computed (a) by numerical integration with a complete dynamical force model and (b) by the approximation in Eq. (6). The difference between these curves is caused primarily by 1- and 2-cycle/rev resonant terms, not accounted for in (6).

1996). The arc is the same 120-minute arc used above, but artificially lowered to 250-km altitude. The signal is computed simply as the gradient of the potential U in (2). As should be expected, a gradiometer is sensitive to relatively short wavelength components of the tidal mass field. In fact, the signal in Figure 3 tends somewhat to

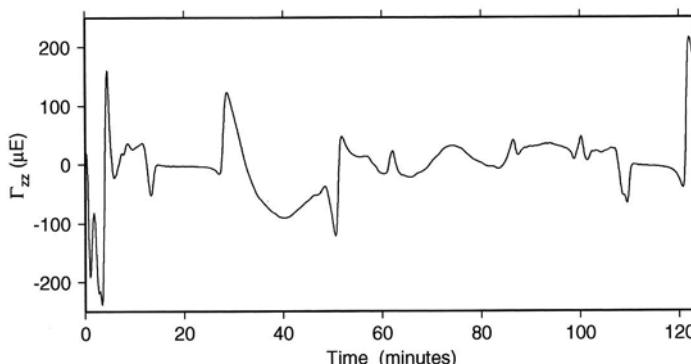


Figure 3. Predicted ocean tide signals in the GOCE gradiometer measurements of $\partial^2 U / \partial z^2$ where U is given by Eq. (2), taken to degree and order 150 for the four largest tidal constituents. The arc is identical to that used for Figure 2 except artificially lowered to 250 km altitude. Vertical axis units are μE , or 10^{-15} s^{-2} .

mimic the tidal elevation field directly below the satellite. For example, the sharp jump near $t = 25$ m occurs as the satellite path passes from land to a coastal region of large tide. Unlike the case for GRACE, the tidal signals in GOCE are more than an order of magnitude smaller than the instrument measurement precision, which for the gradiometer is a few mE/Hz $^{1/2}$ (where 1 Eötvös = 10^{-9} s $^{-2}$). Within the gradiometer measurement band of 0.005–0.1 Hz (ESA, 1999), tidal signals are likely to be detectable only by exploiting their temporal coherence. GPS tracking of GOCE strengthens the low-frequency measurement precision substantially, which further strengthens the tidal sensitivity of the system (see error spectrum below).

We turn now to a discussion of how well these tidal signals can be modeled as an *a priori* correction to the satellite measurements.

3. Errors in Present-Day Tide Models

Comparisons of models with *in situ* tide gauge estimates suggest that open ocean tides are now known to better than 2 cm rms (e.g., Shum et al., 1997; Ray, 1999). Errors in coastal regions, where tides have large amplitudes and high wavenumbers, are probably an order of magnitude worse. Errors are also substantially larger in polar regions above the $\pm 66^\circ$ latitude limits of Topex/Poseidon data. We require here a more definite error model, and we consider two types. First, we simply difference some of the better global tide models now available. Since these models were all developed from Topex/Poseidon measurements, the resulting error model may be somewhat optimistic. Secondly, we adopt the global inverse error estimates described by Egbert and Erofeeva (2002). These are based on a Monte Carlo method that is somewhat dependent on assumed prior error covariances for the altimeter data and for the dynamical tidal equations to which the altimeter data are fit. Employing both types of error models is therefore a useful test of consistency. We concentrate on the M₂ constituent; errors in other tidal constituents are generally smaller, with the possible exception of S₂ errors in low latitudes where contamination from the S₂ atmospheric tide is apparent in some tide models at the level of 1 cm. The error models are described in more detail in Ray et al. (2001).

Figure 4 shows the geoid amplitude spectrum of the M₂ tide and our two different error estimates. These spectra are computed from the complex Stokes coefficients $\delta C_{nm}(t)$ (or from their estimated errors) following Eq. (4). The degree amplitude spectrum in the geoid is

$$\delta N_n = a \left(\sum_{m=-n}^n |\overline{\delta C}_{nm}|^2 \right)^{1/2}, \quad (7)$$

where the overbar denotes the rms over a tidal cycle (12.4 h for M₂) in each complex component of $\delta C_{nm}(t)$. This form of δN_n is consistent with Eq. (4) of Ray et al. (2001) and is also consistent with the geoid spectra of Wahr et al. (1998).

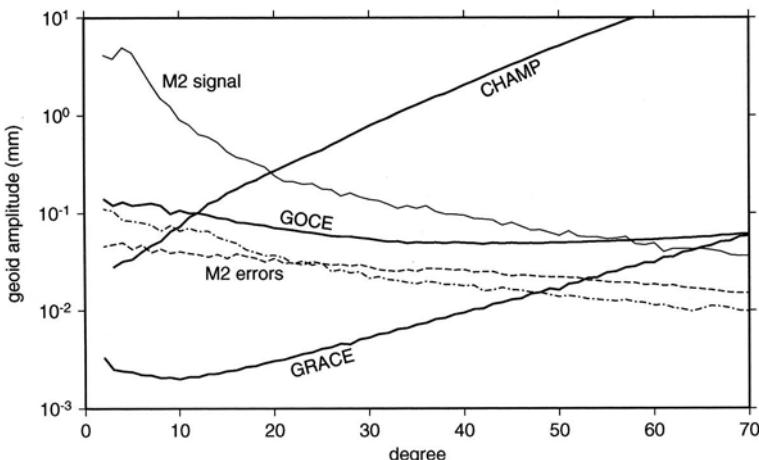


Figure 4. Degree amplitude geoid spectrum of the M₂ tide (thin solid line) and two estimates of its error (thin broken lines, the inverse model being the lower curve at low degree). These curves represent rms spectra over a single tidal cycle (12.4 h). Heavy solid lines denote anticipated sensitivities of three multi-year space geodetic missions, as estimated by Visser et al. (2002). Tidal constituents other than M₂ would fall below these M₂ curves.

In Figure 4 the M₂ error curves are the broken lines, one being the Monte Carlo inverse model, the other being the mean of the differences among models TPXO.5, GOT00.2, and NAO99 (the latter from Matsumoto et al., 2000). The inverse error is the lower curve at small degree n and the higher at large n . The two error curves are reasonably consistent and generally within a factor of 2. The bold lines in Figure 4 represent the sensitivities of three different space geodetic missions after multiple years of data. As expected, the M₂ signal is well above the sensitivities of all three satellite missions, which suggests that high-degree tidal models, perhaps to at least degree 70, are required for GRACE and GOCE. Figure 4 also shows that present-day M₂ errors exceed the GRACE sensitivity below degree 50 or so. The errors also exceed the CHAMP sensitivity below degree 8 or 10. They do not, however, exceed the expected GOCE sensitivity anywhere. This, and the fact that tidal errors are even smaller than the already very small gradient signals shown in Figure 3, suggests that tide model errors are unlikely to be a major concern for the GOCE mission. It is possible that some large tidal errors in a few coastal regions might degrade GOCE analyses, but such errors are far more problematic for satellite altimetry than for satellite gradiometry.

While Figure 4 suggests that tide models must be improved for (or by) the GRACE and CHAMP missions, the temporal mismatch between the tidal and the satellite curves (12.4 h vs. several years) complicates matters. The manifestation of tidal error in GRACE, for example, depends on issues of sampling and on the desired estimation timescales (e.g., instantaneous gravity anomalies, monthly means, or multi-year means).

4. Tidal Errors and GRACE

The effects of tide model errors on GRACE's intersatellite measurements are sufficiently complex that they are difficult to predict without extensive simulations. We show one example below. Nonetheless, some insight can likely be gained from simple considerations of tidal aliasing. GRACE signals such as those in Figure 2b will be aliased in ways quite similar to the aliasing in satellite altimetry. One must bear in mind, however, that the spectrum of tide-induced signals in GRACE is richer than Figure 2b suggests (e.g., Cheng, 2002). GRACE is no altimeter, and some of its tidal errors are entangled with orbit errors with potentially complicated aliasing.

4.1. TIDAL ALIASING

The tidal phases observed by an orbiting spacecraft are determined by the angle between the satellite orbit plane and the tide-raising body. The orbit plane of any high-inclination spacecraft like GRACE precesses so slowly with respect to the sun that solar tides are necessarily aliased into relatively long periods—e.g., semi-annual, annual, or longer. Such periods are of prime importance to climatological studies, making aliased errors at these periods very problematic. (The precession rate with respect to the moon is rapid, and aliasing of lunar tides is rarely problematic unless one happens to be unlucky.) Because the solar tides that GRACE observes will be aliased to long periods, we may predict that errors will be prevalent in the GRACE measurements at these same periods, no matter the exact manner in which the errors are manifested through dynamical complications in the orbit.

From the satellite precession rate we may determine the tidal alias periods. The alias period of the S_1 tide is identical to the satellite nodal period with respect to the mean sun; the S_2 period is half that. The alias period of K_1 is identical to the satellite nodal period with respect to the mean equinox; K_2 is half that. The P_1 alias is identical to the nodal period with respect to a fictitious body moving at twice the sun's mean motion. As of this writing (a few months after launch) the GRACE mean Keplerian elements are $(a, e, I) = (6860.4 \text{ km}, 0.00256, 89.02^\circ)$, implying a node rate of approximately $-0.132^\circ/\text{day}$, or $-1.117^\circ/\text{day}$ with respect to the mean sun. The corresponding alias periods for the primary solar tides are:

K_2	3.74 years	K_1	7.48 years
S_2	161 days	S_1	322 days
P_1	171 days		

Some tidal contamination should therefore be expected at the important semi-annual period and at the interannual periods of 3.7 and 7.5 years. The S_1 alias is also somewhat problematic, since both S_1 (being primarily atmospheric) and the nearby annual cycle (S_a) have strong temporal variability, causing potential overlap of their rather broad spectral peaks.

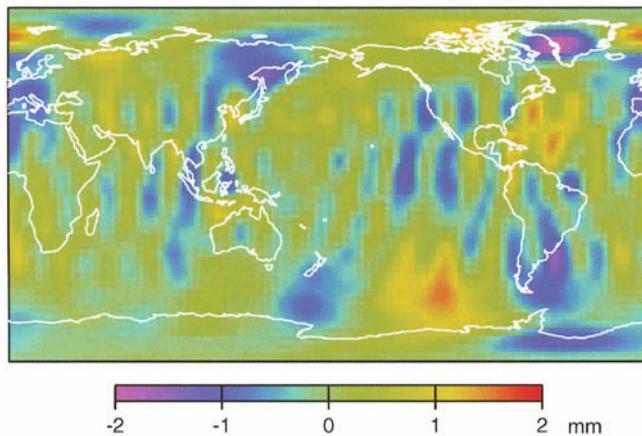


Figure 5. Results of GRACE simulation showing errors in an estimated monthly geoid (in mm) caused by tide model errors. Geoid is evaluated to degree and order 22; higher degree expansions show larger tidal contamination. Note errors are not confined to the oceans, primarily because of the long-period dynamical perturbations to GRACE's range rate.

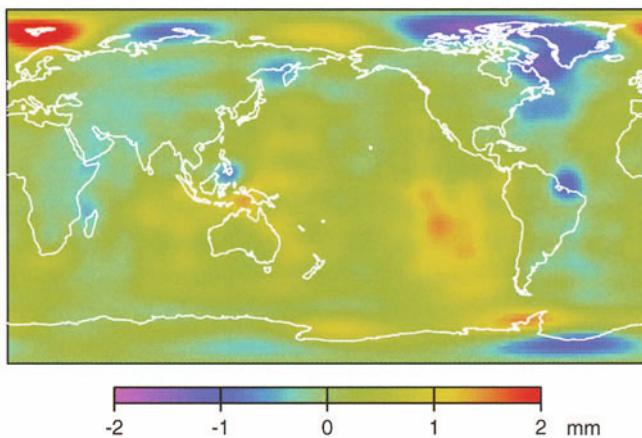


Figure 6. As in Figure 5 except using the simplistic (non-dynamic) approach of Eq. (6). Errors are more localized to the ocean.

4.2. SIMULATION OF MONTHLY GRAVITY ERRORS

Because the GRACE project intends to generate monthly mean estimates of the geopotential (Tapley and Reigber, 2000), it is of interest to anticipate possible contamination of this product caused by errors in the tide model adopted for processing the GRACE measurements. We have estimated this effect by simulating one (arbitrarily chosen) month of GRACE range-rate measurements, using the methods described by Rowlands et al. (2002). The simulation consists of an inversion for a degree-120 static geopotential in the presence of realistic errors in intersatellite

range measurements and in GPS tracking, with and without tide model errors. The latter consists of the difference between tide models GOT00.2 and TPXO.5, to degree 60, for constituents M_2 , S_2 , N_2 , K_2 , O_1 and K_1 . The resulting tidal contamination in the monthly mean geoid estimate is shown in Figure 5. Because errors in monthly geoid estimates rise rapidly with degree, the geoid in Figure 5 is evaluated only to degree 22. Summing to degree 50 results in fairly similar characteristics, although with more orbit striping and roughly double amplitudes.

The simulation represented by Figure 5 is computationally expensive, and it is therefore of interest to see how well the simplistic (and inexpensive) formula (6) can represent the same error field (akin to comparing Figures 2a and 2b). This involves no ephemeris computations and only a relatively small normal matrix. The result is shown in Figure 6. Because this method ignores the dynamical long-period effects, there is no track-type striping and errors are confined primarily to ocean regions. The errors in both figures are of comparable magnitude and they align in a few locations (e.g., Amazon mouth, Weddell Sea), but the method underlying Figure 6 is evidently too simplistic to map the error in detail.

Several points about Figure 5 are worth noting. The presence of geoid errors over both land and ocean is caused by the rich spectrum that tidal forcing induces in the range-rate measurements. Localized errors in ocean tide models can induce global errors in gravity estimation. Secondly, the amplitude of these geoid errors is significant for some anticipated GRACE applications. In particular, any interpretation of the estimated gravity in terms of oceanic mass motions is considerably confounded. Wahr et al. (1998; their Plate 1) estimate annual variations in the geoid of about 1 mm and less, based on a general circulation model, and such signals are comparable to the errors in Figure 5. Thirdly, the errors in Figure 5 are one monthly manifestation and we have no clue to their temporal variations without extending the simulations over a considerably longer timespan. From our discussion of aliasing, however, we anticipate that some of these errors have periods of semi-annual and longer. The S_2 and K_2 tides are likely to be most problematic for monthly mean gravity estimates (Knudsen and Andersen, 2002).

5. Summary Discussion

The ocean tides induce variations in the geoid of several cm, well within the anticipated precisions of the latest space geodetic missions. Yet it is important to realize that the geoid errors in present-day gravity models are a few decimeters over the ocean (Lemoine et al., 1998), whereas errors in the prediction of tidal geoid variations are orders of magnitude smaller. We conclude that missions like GRACE and GOCE can improve our knowledge of the geopotential by an order of magnitude, or more, even with present ocean tide models. Tide model errors become crucial with the extreme precisions required for GRACE to detect temporal gravity changes of the sort discussed by Wahr et al. (1998).

To the extent that Figure 4 is realistic, GRACE itself offers opportunities to improve the long wavelength components of ocean tide models, specifically coefficients at degrees 50 and below, or wavelengths of about 1000 km and longer. CHAMP may also help improve a few low-degree coefficients. These improvements are likely to occur only after multi-year time series have been obtained, because many tides are aliased to nearby frequencies. The solar tides will be especially problematic. For example, GRACE will not sample a complete phase cycle of K_1 until nearly four years have passed, and such long periods will contain substantial power from non-tidal signals. The aliasing of certain tides into semi-annual and near-annual periods will also make solar tide improvements difficult. Lunar tides will be less problematic, and improvements even to M_2 will be most welcome.

Space limitations have forced us to concentrate on the subject of tides as a correction to new satellite gravimetry missions. But if history is a guide, then the markedly enhanced precisions of these new measurements will also yield new information about the tides—solid and atmospheric as well as oceanic—and such improvements can in turn yield further information about other aspects of the Earth system. As just one example, improvements in degree-2 and degree-3 tides shed light on the Earth's anelasticity at frequencies well outside the seismic band. Determination of these low-degree coefficients should improve significantly with the new satellite data.

Acknowledgements

We thank Nikos Pavlis, Ernst Schrama, and John Wahr for useful discussions. Pieter Visser kindly provided the three satellite error curves shown in Figure 4.

References

- Cheng, M. K.: 2002, 'Gravitational perturbation theory for intersatellite tracking', *J. Geod.* **76**, 169–185.
- Egbert, G. D. and S. Y. Erofeeva: 2002, 'Efficient inverse modeling of barotropic ocean tides', *J. Atmos. Oceanic Tech.* **19**, 183–204.
- ESA: 1999, *Gravity Field and Steady-State Ocean Circulation Mission*, Rep. SP-1233, European Space Agency, Noordwijk, 217 pp.
- Knudsen, P. and O. Andersen: 2002, 'Correcting GRACE gravity fields for ocean tide effects', *Geophys. Res. Lett.* **29** (8), 19-1–19-4.
- Lambeck, K.: 1988, *Geophysical Geodesy*, Clarendon Press, Oxford.
- Lemoine, F. et al.: 1998, The development of the joint NASA GSFC and NIMA geopotential model EGM96, NASA Tech. Memo. 206861, Goddard Space Flight Center, Greenbelt, 575 pp.
- Matsumoto, K., T. Takanezawa, M. Ooe: 2000, 'Ocean tide models developed by assimilating Topex/Poseidon altimeter data into hydrodynamical model', *J. Oceanogr.* **56**, 567–581.
- Pavlis, E. C. and L. Iorio: 2002, 'The impact of tidal errors on the determination of the Lense-Thirring effect from satellite laser ranging', *Int. J. Mod. Phys. D*, **11**, 599–618.

- Ray, R. D.: 1999, A global ocean tide model from Topex/Poseidon altimetry, NASA Tech. Memo. 209478, Goddard Space Flight Center, Greenbelt, 58 pp.
- Ray, R. D., R. J. Eanes, G. D. Egbert, N. K. Pavlis: 2001, 'Error spectrum for the global M_2 ocean tide', *Geophys. Res. Lett.* **28**, 21–24.
- Rowlands, D. D., R. D. Ray, D. S. Chinn, F. G. Lemoine: 2002, 'Short-arc analysis of intersatellite tracking data in a gravity mapping mission', *J. Geod.* **76**, 307–316.
- Schrama, E. J. O.: 1996, 'Gravity research missions reviewed in light of the indirect ocean tide potential', in R. H. Rapp, A. A. Cazenave and R. S. Nerem (eds.), *Global Gravity Field and Its Temporal Variations*, Springer, Berlin, pp. 131–140.
- Shum, C. K. et al.: 1997, 'Accuracy assessments of recent global ocean tide models', *J. Geophys. Res.* **102**, 25173–25194.
- Tapley, B. D. and C. Reigber: 2000, 'The GRACE mission: status and future plans' (abstract), *EOS, Trans. AGU*, **81**, Fall Suppl., F307.
- Visser, P. et al.: 2002, 'The European Earth Explorer Mission GOCE: Impact for the geosciences', in J. Mitrovica and L. Vermeersen (eds.), *Ice Sheets, Sea Level and the Dynamic Earth*, American Geophysical Union, Washington.
- 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', *J. Geophys. Res.* **103**, 30205–30229.

Address for Offprints: R. D. Ray, NASA GSFC, Code 926, Greenbelt MD 20771, USA
richard.ray@gsfc.nasa.gov

THE ELUSIVE STATIONARY GEOID

MARTIN VERMEER

Institute of Geodesy, Department of Surveying, Helsinki University of Technology

Received: 10 May 2002; Accepted in final form: 21 January 2003

Abstract. We discuss the various problems occurring when trying to fix a geoid or geopotential model using sea level observations sampled during a limited time span from a bounded geographical domain. Such problems are on the one hand aliasing and spectral leakage, and on the other, the non-conservation of matter over only part of the world ocean.

In the light of these issues we discuss whether it is sensible to include in a definition of the global geoid the radially symmetric part of either the mean sea level field itself, or its linear or nonlinear time dependence, arriving at a negative conclusion.

Keywords: Stationary geoid, mean sea level, aliasing, spectral leakage.

1. Introduction

It would be tempting, and quite feasible (Vermeer and Poutanen, 1996), to write a paper on how to define the geoid, and then on how to define, and determine, the stationary geoid. Such a paper would address the question how the standard atmosphere should be treated such that the same geopotential model can serve both gravimetrists and geodesists on the Earth surface and in space; it would discuss the proper way to take into account the permanent part of the tidal potential, so as to allow the geoid obtained to be useful both for geodesists and their vertical reference systems on land, and for oceanographers and climatologists studying ocean currents.

However, I have chosen to concentrate on the more concrete problem of *data sampling*, both in space, and (especially) in time.

2. About Stationarity and Trends

Stationarity can be defined by describing the field in question as

$$f(\varphi, \lambda, t) = a(\varphi, \lambda) + b(\varphi, \lambda)t + c(\varphi, \lambda, t), \quad (1)$$

where

$$a(\varphi, \lambda) + b(\varphi, \lambda)t$$



represents the *stationary part*, and $c(\varphi, \lambda, t)$ the time-varying part. This partition is usually done in such a way as to make c “as small as possible” in some norm sense, e.g., the square integral sense over the time and space domain of study.

Looking at the above equation, the phenomenon described can be anything: not only the geoid, but many other things, like ocean surface, air pressure, mean temperature, infrared luminosity – whatever. In the case of atmospheric conditions, we would call a, b “climate” and c “weather”. And b , the interesting term, would represent climate change.

The question arises, how well can we determine a and b ? We will have available observational data from a limited time span (t_0, t_1) . We want to estimate values for the functions $a(\varphi, \lambda)$ and $b(\varphi, \lambda)$, which are as “good” as possible, i.e., as close to “physical reality” as possible. But physical realism in this sense is somewhat “metaphysical”.

Therefore, it is better to require that the estimates (a, b) to be obtained are as *long lasting* as possible, i.e., the values obtained from a limited time series will “stand up” as well as possible against longer time series becoming available. This is a measure of “quality” that will be generally acceptable, despite its somewhat theoretical background.

A straightforward use of an existing observation time series of a phenomenon, $f_i = f(t_i)$, $i = 1, \dots, n$, is to perform a *linear regression* on the observations. A representation $f(t) = a + b(t)$ can be found in this way, which is optimal in the statistical sense – provided the statistical properties of the phenomenon’s observation process are correctly modelled by

$$f(t_i) = a + bt_i + \underline{n}_i,$$

where \underline{n}_i is *uncorrelated noise*, i.e., $E(\underline{n}_i) = 0$ and $E(\underline{n}_i \underline{n}_j) = 0$ for $i \neq j$. This is usually not the case for real physical processes! A more realistic model is

$$\underline{f}(t_i) = a + bt_i + \underline{s}(t_i) + \underline{n}_i,$$

where $\underline{s}(t)$ represents a *signal function*, one part of the physical process we are trying to study. Simply applying linear regression will not produce the right a and b , but, depending on the nature of the function \underline{s} , some of this function will get mixed in too. This usually is referred to as *aliasing*.

A formally correct way of handling this situation, but requiring information on the statistical properties of both \underline{n} and \underline{s} , is *least squares collocation with parameters* as treated in the geodetic literature, e.g., (Mikhail and Ackermann, 1976) or (Vaněček and Krakiwsky, 1986). Basically, \underline{s} is just treated as another observational noise term, to be properly modelled by an appropriate* variance-covariance matrix in the estimation of a and b .

* It is in particular critical to account properly for the *correlations* in time within $\underline{s}(t)$, which in the frequency domain appear as lower frequency constituents, a “coloured noise” signature.

2.1. DERIVING TRENDS FROM “DIRTY” TIME SERIES

2.1.1. *Experiences Processing Tide Gauge Data*

The author acquired some experience with the derivation of trends from “dirty” (i.e., containing a coloured signal constituent) time series, when in the 1980’s processing tide gauge observations from the Finnish coast, together with personnel from the Finnish Institute of Marine Research (Vermeer et al., 1988), in order to get a handle both on the postglacial land uplift in the thirteen Finnish tide gauge locations and on the secular behaviour of the Baltic Sea level.

The publication is of some interest even today, because the attempt was made to estimate in a common computation tide gauge reference constant and uplift values as well as mean sea level variations, according to the observation equation:

$$h_i(t) = a(t) + b(t)(\varphi_i - \varphi_0) + c(t)(\lambda_i - \lambda_0) + z_i + u_i(t - t_0) + \varepsilon_i(t),$$

where

- $a(t)$, $b(t)$, $c(t)$ are bias and tilt coefficient time series of a planar fit through simultaneous tide gauge readings, which describe the spatially linear time dependent behaviour of sea level;
- z_i , u_i are time gauge constant offset and trend (“land uplift”) constants for tide gauge i ;
- φ_0 , λ_0 are the reference latitude and longitude;
- t_0 is a reference time near the middle of the interval of study.

For the huge number of time epochs t , brute force solution of this problem would have been numerically challenging at the time.

In the publication mentioned, a separate study was made of the effect of the cut-off year on the resulting tide gauge land uplift. It was concluded that

the choice of data interval has a significant influence on the results obtained, especially on the bias term affecting all tide gauges equally.

This can be seen in the following table of computed land uplift figures u (unit: mm/year) for tide gauge Hanko, having the longest time series (starting in 1887):

1980	1981	1982	1983	1984	1985	80 – 85
3.37	3.16	3.09	2.82	2.90	2.90	+0.47

The effect is impressive, it dominates any observational accuracy effects! This is something to be kept in mind when judging the results of time series analyses done by “brute-force” linear regression techniques.

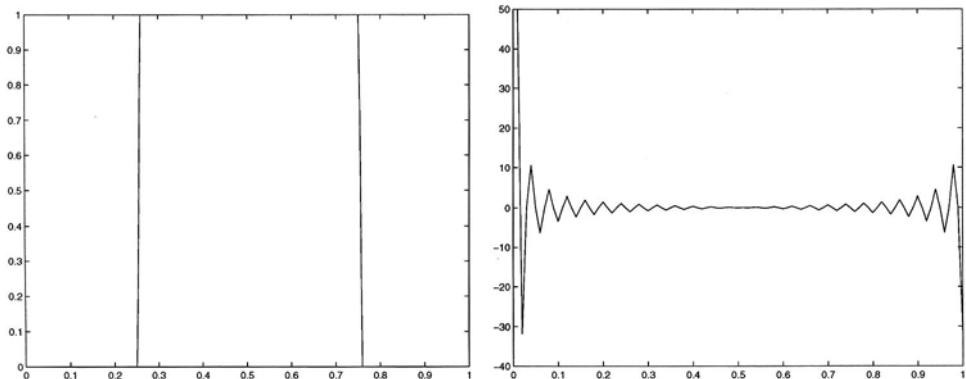


Figure 1. The original box function and its Fourier transform

2.1.2. The Effect of the “Box Filter”

A good practical technique to get around some of these problems is referred to as *tapering*. It is based on the understanding that the effect is due to the spectral content of a rectangular box filter,

$$B(t) \equiv \begin{cases} 1 & \text{for } t \in (t_0, t_1) , \\ 0 & \text{for } t \notin (t_0, t_1) , \end{cases} \quad (2)$$

leaking into the resulting trend found (and into other spectral constituents). The core of the problem are the pronounced discontinuities in this function.

The Fourier transform of a box is

$$\begin{aligned} \mathcal{F}\{B\}(\omega) &= \int_{-\infty}^{+\infty} B(t) e^{2\pi i t \omega} dt = \\ &= \int_{t_0}^{t_1} e^{2\pi i t \omega} dt = \left[\frac{1}{2\pi i \omega} e^{2\pi i \omega t} \right]_{t_0}^{t_1} = \\ &= \frac{e^{2\pi i \omega \bar{t}}}{2\pi i \omega} (e^{2\pi i \omega \Delta t} - e^{-2\pi i \omega \Delta t}) = \\ &= \frac{e^{2\pi i \omega \bar{t}}}{\pi \omega} \sin 2\pi \omega \Delta t, \end{aligned}$$

where $\bar{t} \equiv \frac{1}{2}(t_0 + t_1)$ and $\Delta t \equiv \frac{1}{2}(t_1 - t_0)$. This produces an oscillating pattern in frequency space, with the amplitude envelope described by the function $\frac{1}{\pi \omega}$. See Fig. 1, which however depicts a discrete transform on 100 points. For $\bar{t} = 0$ we have a real-valued function, $\frac{\sin 2\pi \omega \Delta t}{\pi \omega}$, which is also the amplitude function.

As the function g analyzed is actually the product of the box function and the function f of interest, it follows by the convolution theorem, that any frequency present in the function of study will be “smeared out” into the zero frequency

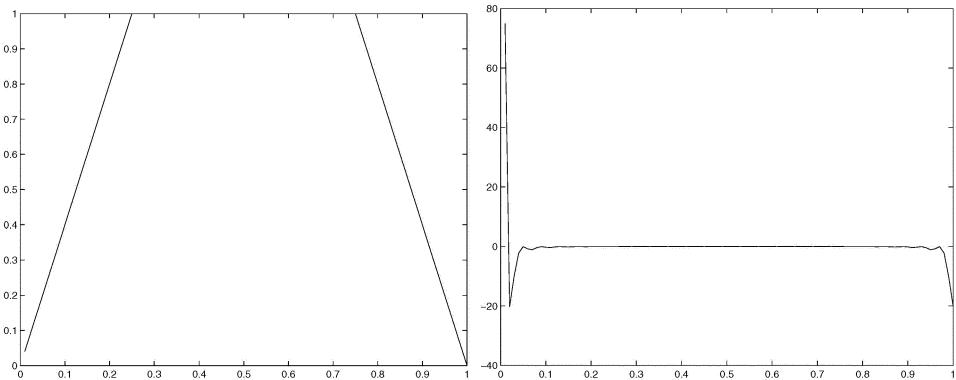


Figure 2. Linear taper and Fourier transform

component in proportion to this coefficient:

$$\begin{aligned}\mathcal{F}\{g\}(\omega) &= \int_{-\infty}^{+\infty} \mathcal{F}\{f\}(\omega') \cdot \mathcal{F}\{B\}(\omega - \omega') d\omega' \Rightarrow \\ \mathcal{F}\{g\}(0) &= \int_{-\infty}^{+\infty} \mathcal{F}\{f\}(\omega') \cdot \mathcal{F}\{B\}(-\omega') d\omega'.\end{aligned}$$

2.1.3. Using a Linear Taper

The above mentioned oscillation effect can be significantly reduced by choosing, instead of a block function, a linear taper, which is a continuous function of t :

$$L(t) = \begin{cases} t/\Delta t & \text{for } t \in (\bar{t} - 2\Delta t, \bar{t} - \Delta t) \\ 1 & \text{for } t \in (\bar{t} - \Delta t, \bar{t} + \Delta t) \\ 2 - (t - \bar{t})/\Delta t & \text{for } t \in (\bar{t} + \Delta t, \bar{t} + 2\Delta t) \\ 0 & \text{elsewhere.} \end{cases}$$

Figure 2 (left) illustrates $L(t)$. It must be obvious that, in the presence of low to medium frequency signal in the original function $f(t)$, the box filter – i.e., a sharp cut-off of the data at the edges of the time interval (t_0, t_1) of study – will lead to a much worse contamination of the zero frequency component of the function $g(t)$ actually used, than if we use a “soft” cut-off, using a linear or an even higher order taper.

To derive the Fourier transform for the linear taper, note that

$$L(t) = T_{-\Delta t}(t) + T_0(t) + T_{\Delta t}(t),$$

where T is the triangle function

$$T_0(t) = \begin{cases} 1 + t/\Delta t & \text{for } -\Delta t < t < 0 \\ 1 - t/\Delta t & \text{for } 0 < t < \Delta t \\ 0 & \text{elsewhere,} \end{cases}$$

and where $T_{t_0}(t)$ is obtained from $T_0(t)$ by a shift: $t \rightarrow t - t_0$.

The Fourier transform of T_0 is (leaving out details):

$$\begin{aligned}\mathcal{F}\{T_0\}(\omega) &= \int_{-\infty}^{+\infty} T_0(t) e^{2\pi i \omega t} dt = \\ &= \int_{-\Delta t}^0 \left(1 + \frac{t}{\Delta t}\right) e^{2\pi i \omega t} dt + \int_0^{+\Delta t} \left(1 - \frac{t}{\Delta t}\right) e^{2\pi i \omega t} dt = \\ &= -\frac{1}{\pi \omega}(0) + \frac{2}{\Delta t} \left(\frac{1}{2\pi \omega}\right)^2 \sin 2\pi \omega \Delta t = \frac{1}{2\pi \omega \Delta t} \frac{\sin 2\pi \omega \Delta t}{\pi \omega}.\end{aligned}$$

For the linear taper, the Fourier transform is the sum of three rotated transforms of T :

$$\begin{aligned}\mathcal{F}\{L\}(\omega) &= \int_{-\infty}^{+\infty} (T_{-\Delta t} + T_0 + T_{\Delta t})(t) e^{2\pi i \omega t} dt = \\ &= (e^{-2\pi i \omega \Delta t} + 1 + e^{+2\pi i \omega \Delta t}) \mathcal{F}\{T_0\}(\omega) = \\ &= (1 + 2 \cos 2\pi \omega \Delta t) \frac{1}{2\pi \omega \Delta t} \frac{\sin \pi \omega \Delta t}{\pi \omega}.\end{aligned}$$

This amplitude function $\frac{1}{2\pi \omega \Delta t} \frac{\sin \pi \omega \Delta t}{\pi \omega}$, looks similar to the box transform, $\frac{\sin 2\pi \omega \Delta t}{\pi \omega}$, but multiplied by $\frac{1}{2\pi \omega \Delta t}$. This is why for $2\pi \omega \Delta t = 1$ (the lowest frequency, corresponding to one wave across the length of the box or filter function) we obtain the result that the two transforms are the same. For $2\pi \omega \Delta t = k$, the ratio will be $1/k$ and it will thus drop off very quickly with increasing frequency.

Obviously these formulas are invalid for $\omega = 0$.

Other smoothing tapers are used in geoid computation in regional areas, such as cosine taper and cubic taper, for which even the function's first derivative is continuous. The gain is, however, marginal (Vermeer, 1992).

3. Spatial Sampling of Global Sea Level

We refer to the “mean sea level” as an average over the entire globe, we obviously mean the value of sea level averaged over the whole globe, or rather, that part of the globe actually covered by ocean.

In practice, however, nowadays often a sub-set of this area is used that is within the orbital inclination reach of *radar altimetric satellites*, like TOPEX/ Poseidon, which has been used to define the mean sea level by its radar altimeter, where the satellite was continually positioned in space by an on-board GPS receiver.

While obviously much, much better than the device of using a seriously location-biased and undersampled ensemble of coastal and island tide gauges, this restriction still is problematic. Sea water is continually in motion, and the total volume of water in a limited area may be highly variable in a broad range of time scales.

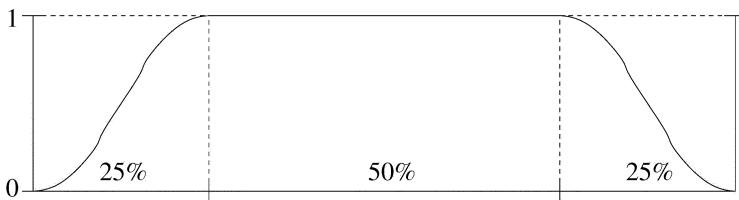


Figure 3. A 50% cubic taper

This aspect matters especially for TOPEX/Poseidon with its orbital inclination of only 66° . Therefore some effort was made to “weave” together the networks of cross-over arcs from other satellites as well (like ERS-1, ERS-2 and GEOSAT). As the ERS satellites have inclinations of $98^\circ.5$, one thus reaches Northern and Southern latitudes of $81^\circ.5$, leaving only small polar caps uncovered.

In the spatial domain of the world ocean, the occurrence of a sharp cut-off at the inclination limit of the satellites used will introduce similar aliasing problems as for the same technique in the time domain. Tapering techniques may provide a simple solution if a formally correct treatment is, e.g., too CPU intensive.

4. The Problem of Non-globality

The *total* volume of water in the world oceans is much less variable than that in parts of the ocean: changing it requires, e.g., large volumes of water to be produced by melting of continental ice sheets or by thermal expansion of existing water masses, both processes of high inertia. These processes, if occurring in a secular way, should be captured by the b term in Eq. (1).

When working with data on a latitude-bounded ocean domain, one therefore must, in addition to the effect of sharp latitude boundaries referred to above, also expect a significant amount of leakage on all time scales caused by water masses arbitrarily crossing these boundaries.

This problem affects mostly the long-wavelength components of the physical field being studied. For the global constant part, removing it altogether (the approach described in Sect. 5, of forcing radially symmetric biases, trends and other temporal variations to zero), obviates the problem.

For other problems than geoid determination, however, this option may not exist.

One can see a theoretical similarity between this the radially symmetric part removal technique and regularization techniques aimed at solving the “polar gap” problem for all geopotential satellite missions not flying in strictly polar orbits.

5. How to Define a Reference Geoid

When studying the *geoid* rather than the physical ocean surface, the above considerations have consequences. We should question the wisdom of too close a coupling between the physical mean sea surface – practically determinable only over a subset of the global ocean, and more noisy over a broad temporal spectrum as a result – and the definition of the geoid.

One can look at this in a slightly different, more physical fashion. If we rewrite Eq. (1) as

$$N(\varphi, \lambda, t) = a_0 + a_1(\varphi, \lambda) + [b_0 + b_1(\varphi, \lambda)]t + [c_0(t) + c_1(\varphi, \lambda, t)], \quad (3)$$

we may now consider

$$N_0(t) \equiv a_0 + b_0 t + c_0(t)$$

as a global, radially symmetric, time dependent *central term*.

Let us ask first what we use a model of the true geoid, or geopotential, for. The following applications have been widely mentioned:

1. Provide a global height reference, by allowing the reduction of GPS-derived geometric heights h to the geoid as a reference surface, obtaining orthometric heights $H \equiv h - N$.
2. Provide a reference for describing variations in ocean surface level, in a dynamically meaningful way for oceanographic current modelling, i.e., deviations from this reference surface should represent true energy differences.

It is typical for both fields of application that we speak of height, but think of – potential – *energy*. Clearly, only *differences* in energy are of interest to us. The absolute level of the equipotential surface that we call “geoid” is of no practical (real) relevance. Only tradition and reverence for C. F. GAUSS let us make the attempt to have the geoid as close as possible to the “mean sea level”.

But, as local mean sea level at the tide gauge(s) used by countries to establish their vertical references will deviate, due to local sea surface topography, by as much as ± 1 m from the global mean anyway*, we are free, from a practical viewpoint, to fix our geoid to any level reasonably close to global mean sea level.

Therefore, we propose to proceed as follows:

Fix the reference potential of the geoid to that of the GRS80 reference ellipsoid (with a correction for the standard atmosphere), a constant.

Note also that the “currently best” value for the Earth’s equatorial radius is $a = 6378136.3$ m, 70 cm shorter than the GRS80 value. The EGM96 global geopotential expansion uses a value 53 cm short of the GRS80 value (Lemoine et al., 1998, pp. 11-3, 11-4).

* Tidal and weather-induced variations in a wide range of time scales and of comparable amplitudes exist as well.

Referring back to Eq. (3) this means that we have

$$N(\varphi, \lambda, t) = a(\varphi, \lambda) + b(\varphi, \lambda)t + c(\varphi, \lambda, t),$$

where we require

$$\iint_{\Sigma} a(\varphi, \lambda) d\sigma = \iint_{\Sigma} b(\varphi, \lambda) d\sigma = 0,$$

where Σ is the whole geocentric unit sphere. Furthermore, we must obviously also require

$$\iint_{\Sigma} c(\varphi, \lambda, t) d\sigma = 0 \quad \forall t \in (t_0, t_1).$$

Under those assumptions a represents the stationary (time constant) geoid, b the geoid trend in time – the greatest part of which is due to postglacial rebound, to redistribution of water between continental ice sheets and ocean, and regional tectonic vertical motions – and c represents all zero-sum mass redistributions in ocean and atmosphere on a wide range of time scales taken together.

Note, that the traditional gravimetric geoid determination technique from gravity anomalies Δg , using the famous STOKES integral

$$N = \frac{R}{4\pi g} \iint_{\Sigma} S(\psi) \Delta g d\sigma,$$

with the Stokes kernel function $S(\psi)$ of the spherical distance between data and target points, imposes the above conditions implicitly.* While it is possible to use gravimetric methods, both on the Earth surface and from space, to get a handle on the central force field term GM_{\oplus} , the choice of a particular combination of GM and the geoid reference potential W_0 will always be a somewhat arbitrary.

6. Conclusions

The separation of the true, instantaneous geoid into a “stationary” and a (non-linearly) time-varying part is difficult to achieve in a physically meaningful and unique sense. One may even question if this can be done in a physically meaningful way.

* In fact, it imposes the further condition that the three degree-1 spherical harmonic constituents vanish as well:

$$\iint_{\Sigma} f(\varphi, \lambda, t) \sin \varphi d\sigma = \iint_{\Sigma} f(\varphi, \lambda, t) \cos \varphi \sin \lambda d\sigma = \iint_{\Sigma} f(\varphi, \lambda, t) \cos \varphi \cos \lambda d\sigma = 0$$

for $f \equiv a, b, c$. A condition that doesn’t make physical sense for other fields, but for the geopotential naturally expresses the choice of the Earth’s centre of mass as the co-ordinate origin.

We therefore propose to fix the globally averaged geoid level to the conventional value for GRS80 and not to allow for time variations. This procedure still allows for global, but zero-centred, geoid maps for the constant part a , for a time trend b and for irregular time variations c to be determined. Actually separating bt from $c(t)$ in a clear-cut way is quite challenging and may not necessarily be meaningful.

References

- Lemoine, F., S. Kenyon, J. Factor, R. Trimmer, N. Pavlis, D. Chinn, C. Cox, S. Klosko, S. Luthke, M. Torrence, Y. Wang, R. Williamson, E. Pavlis, R. Rapp, and T. Olson: 1998, *The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96*. Greenbelt, Maryland 20771: National Aeronautics and Space Administration, Goddard Space Flight Center.
- Mikhail, E. M. and F. Ackermann: 1976, *Observations and Least Squares*. Harper & Row.
- Vaníček, P. and E. Krakiwsky: 1986, *Geodesy – The Concepts*. Amsterdam: Elsevier Science Publishers.
- Vermeer, M., J. Kakkuri, P. Mälkki, K. Kahma, M. Leppäranta, and H. Boman: 1988, 'Land uplift and sea level variability spectrum using fully measured monthly means of tide gauge readings'. In: *Finnish Marine Research*. Helsinki: Finnish Institute of Marine Research, pp. 3–75.
- Vermeer, M. and M. Poutanen: 1996, 'A Modified GRS-80 Normal Field Including Permanent Tide and Atmosphere'. In: J. Segawa, H. Fujimoto, and S. Okubo (eds.): *Proceedings, IAG International Symposium on Gravity, Geoid and Marine Geodesy (GraGeoMar96), Tokyo, Sep. 30 - Oct. 5, 1996*. Tokyo, pp. 515–522.
- Vermeer, M.: 1992, 'FGI Studies on Satellite Gravity Gradiometry. 3. Regional high resolution geopotential recovery in geographical coordinates using a Taylor expansion FFT technique'. Report 92:1, Finnish Geodetic Institute, Helsinki.

Address for Offprints: Martin Vermeer, Institute of Geodesy, Department of Surveying, Helsinki University of Technology, FIN-02015 HUT Finland

GEODETIC METHODS FOR CALIBRATION OF GRACE AND GOCE

JOHANNES BOUMAN AND RADBOUD KOOP

SRON National Institute for Space Research

Sorbonnelaan 2, 3584 CA Utrecht, The Netherlands

Received: 27 May 2002; Accepted in final form: 20 September 2002

Abstract. It is beyond doubt that calibration and validation are essential tools in the process of reaching the goals of gravity missions like GRACE and GOCE and to obtain results of the highest possible quality. Both tools, although general and obvious instruments for any mission, have specific features for gravity missions. Therefore, it is necessary to define exactly what is expected (and what cannot be expected) from calibration and what from validation and how these tools should work in our case. The general calibration and validation schemes for GRACE and GOCE are outlined. Calibration will be linked directly to the instrument and the measurements whereas validation will be linked to data derived from the original measurements. Calibration includes on-ground, internal, and external calibration as well as error assessment. The calibration phase results in corrected measurements along with an a posteriori error model. Validation of e.g. calibrated measurements or geoid heights means checking against independent data to assess whether there are no systematic errors left and/or whether the error model describes the true error reasonably well. Geodetic methods for calibration typically refer to external calibration and error assessment, and will be illustrated with an example.

Keywords: GRACE, GOCE, calibration, validation

1. Introduction

The gravity missions GRACE and GOCE will deliver high resolution, high accuracy static and temporal gravity field information (Tapley and Reigber, 1999; ESA, 1999). Before this gravity field information can be derived from the observations it is important that systematic errors have been removed to the extent possible from the data and that the data accuracy has been assessed. When the gravity field information has been extracted from the data, it is necessary to check whether there are no undetected systematic errors left. The correction for systematic errors, the error assessment and the final quality check are generally referred to as cal/val (calibration and validation) and are the topic of this paper.

We will first describe more carefully how to define calibration and validation is the case of gravity missions in Section 2. We will then focus on calibration for GRACE and GOCE in Section 3. In Section 4 we will discuss external calibration in more detail, in particular for GOCE satellite gravity gradient (SGG) observations, and we will give an example in Section 5.



Space Science Reviews **108:** 293–303, 2003.

© 2003 Kluwer Academic Publishers.

2. Calibration and Validation

The aim of calibration is to achieve “correct” observations, i.e., observations from which to our best knowledge any error from the measurement process has been removed, and to provide information concerning the measurement error. The former is also called signal calibration, while the latter is called error assessment. In the signal calibration step one wishes to correct the observations for measurement or instrument errors. To do so, first instrumental or mathematical calibration methods and tests are performed. These allow it to determine calibration parameters such as scale factors, biases, etc. Then the calibration parameters are applied to the observations (the actual correction), leading to the “correct” observations. In the error assessment step the aim is to establish a proper realistic error model of the observations. This can be done by updating or verifying some a priori formal error model based on the analysis of the observations or on external data or knowledge. The transformation of formal errors into realistic errors is also referred to as error calibration. In general, calibration deals with imperfections of the measurement process, which are due to instrument errors and imperfections, and pre-processing errors. For our purposes it is appropriate to define calibration as *the procedure to determine parameters and the application of these parameters to the instrument read-outs in order to obtain quantities in the required physical units with known accuracy* (Koop et al., 2001).

It is the aim of validation to provide information about the success of the calibration and to assess the quality of end products. Such accuracy assessment may be based on internal evaluation (inter-comparison of parallel solutions from the same data) and external evaluation (comparison with existing gravity field knowledge, geophysical models or data from other measurement campaigns), as well as the assessment of the spatial and spectral information content of the solution. A comparison of the solution with gravity “standards” may establish the “absolute” information content. In general, validation gives a quality report of the solution but it does not provide corrections to the data. The latter is the key issue which makes up the difference between calibration and validation as defined here. For our purpose validation may thus be defined as *the application of methods to compare data products derived from the original measurements with existing independent data or knowledge in order to assess the quality of these data products and to make sure that the measurement process, error estimation and the calibration have been performed well* (Koop et al., 2001). For the gravity missions GRACE and GOCE it is important to decide whether the quality of existing static and temporal gravity field data and geophysical models will be good enough for validation purposes at all spatial and temporal scales. Actually, these missions are intended to greatly improve or even “set the standards” of our global gravity field knowledge. Therefore the demands on the quality of the “validating data” will indeed be high. We will not pursue the topic of validation here any further, but focus on calibration in the following sections.

3. Calibration for GRACE and GOCE

3.1. GENERAL

When discussing cal/val it is convenient to refer to the different data product levels which are defined for these kind of satellite data. Four data levels can be distinguished. Level 0 are the raw telemetry data as delivered by the instruments and the satellite, level 1 are the pre-processed measurements (e.g., calibrated and corrected SGG observations), and level 2 are the geodetic products derived from the level 1 data (gravity field models, for example, and precise orbits). Further use of level 2 data in scientific fields such as oceanography and geodynamics results in level 3 data products. Calibration in the sense as it is described in Section 2 involves level 1 data, whereas validation deals with level 2 data.

Calibration of the measurements, that is the determination and application of corrections, may be regarded as a three-step procedure. The first step is the pre-launch *on-ground calibration*. In this step the instrument scale factors are determined as good as possible and the instruments are aligned with each other and with the satellite as good as possible. A second calibration is needed when the satellite is in its orbit. This is called *in-flight* or *internal calibration*. Internal calibration is necessary because the launch may have disturbed the instruments or because proper calibration on ground is impractical, too expensive, or not possible. In the case of the accelerometers the 1-g-environment causes the on-ground calibration to be too inaccurate. Both, GRACE and GOCE, will use a known thruster signal to internally calibrate the accelerometers, and the K-band ranging (KBR) alignment in case of GRACE. For both GRACE and GOCE, on-ground calibration is a unique event, whereas the internal calibration may be repeated in the case of unacceptable instability of the calibration parameters. For GOCE the latter may be needed once a month at most, for GRACE maybe every 90 days, but longer intervals between subsequent calibration procedures are desirable to obtain a continuous data stream.

Finally, a third calibration is needed which is called *external calibration*. In this step, the internally calibrated measurements are tested against independent data (hence external) with the aim to check whether the internal calibration worked well and to determine absolute accelerometer scale factors (e.g.) which cannot be determined with sufficient accuracy in the earlier calibration phases. Because external data is used in this third calibration step, external calibration is often confused with validation, since the latter also makes use of external data. But, as said before, calibration (and also external calibration) implies the active correction of the observations, which is absent in the case of validation. Note that the results of the validation step (consisting, e.g., of a quality report) may lead to a feedback to the calibration procedure, for example if certain systematic errors appear to be still present in the end products, and that certain calibration steps might need a reassessment and improvement. In that sense there is a strong relation between (external)

calibration and validation. For brevity and clarity such feedback mechanisms will not be further discussed here.

3.2. GRACE CALIBRATION

Before treating in some more detail the calibration of GOCE SGG data, let us address briefly the GRACE calibration. The on-ground calibration for GRACE consists among others of the determination of accelerometer scale factors and the alignment of the accelerometers and the star cameras with the satellites. With the in-flight internal calibration one tries to physically adjust the position of the accelerometer in such a way to position it at the centre of mass (CoM) of the satellite. Furthermore, updates of scale factors and biases of the accelerometers are determined and the alignment of the KBR boresight with the star camera is determined. Finally, several tests are performed, like orbit computation based on SLR (Satellite Laser Ranging) and GPS data, comparison of the KBR residuals with relative satellite positions from GPS and the so-called “null test” of gravity variability in “quiet” regions (Egypt and Saudi Arabia) (see (Velicogna et al., 2000)), and comparison of temporal gravity signals with ocean bottom pressure recordings. Within the context of GRACE these tests are referred to as “external calibration”, although in the sense of our definitions in Section 2 these tests should be characterized as validations, since they do not lead to corrections of the observations in a direct manner. The major concern with GRACE calibration is aliasing of short period gravity field variations into the 30 day solutions. Furthermore, the spatial coverage of calibration areas is limited. The question remains whether calibration parameters deduced from such tests are applicable to the global data set.

3.3. GOCE CALIBRATION

The gravity gradiometer of GOCE consists of six 3-axes accelerometers, see Figure 1. The X -axis is, ideally, in the flight direction of the satellite, the Z -axis is almost radially outwards in the plane spanned by the position vector of the satellite in an Earth-fixed frame and the velocity vector, and the Y -axis completes the right-handed frame. This $\{X, Y, Z\}$ -frame is the so-called Local Orbital Reference Frame (LORF), (Cesare, 2002).

In each accelerometer a proof mass is electrostatically suspended and actively controlled at the centre of a cage by means of a feedback mechanism. The proof mass displacement relative to the electrodes is measured by capacitive sensors. The control voltages are representative of the accelerations of the proof mass relative to the cage. The acceleration $\vec{a}_{i\alpha}$, $i = 1, 2$, $\alpha = X, Y, Z$ of the proof mass with respect to the cage is the sum of the accelerations due to: the non-gravitational forces, the gravity gradient tensor, other satellite masses (satellite self-gravity), the coupling of the proof mass with the external magnetic field, the satellite angular rotation and acceleration, and the Coriolis acceleration (Cesare, 2002; ESA, 1999). If the accelerometer would be in the CoM then the accelerometer would only sense

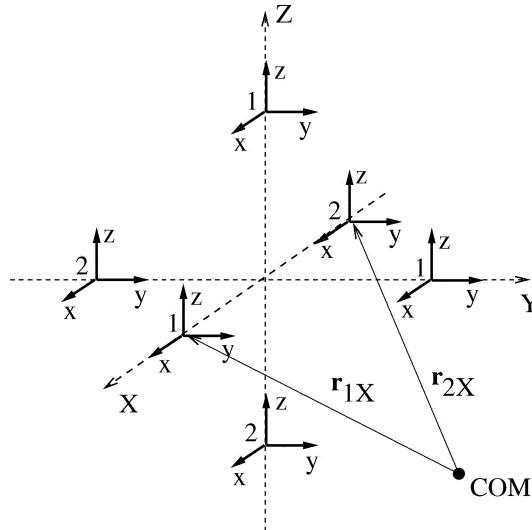


Figure 1. The six accelerometers in the Local Orbital Reference Frame.

the non-conservative forces, the self-gravity, and magnetic field forces. This is the GRACE concept where the non-conservative forces are to be measured.

The common mode and differential mode accelerations are

$$\begin{pmatrix} \vec{a}_{c\alpha} \\ \vec{a}_{d\alpha} \end{pmatrix} := \frac{1}{2} \begin{pmatrix} \vec{a}_{1\alpha} + \vec{a}_{2\alpha} \\ \vec{a}_{1\alpha} - \vec{a}_{2\alpha} \end{pmatrix}, \quad \alpha = X, Y, Z. \quad (1)$$

Thus the eighteen observed accelerations in $\vec{a}_{i\alpha}$ are transformed to three common mode accelerations $\vec{a}_{c\alpha}$ (= nine observations) and three differential mode accelerations $\vec{a}_{d\alpha}$ (= nine observations). If it is assumed that the centre of the frame formed by the six accelerometers is in the CoM, and if the self-gravity and magnetic field forces are neglected, then the common mode (CM) is proportional to the non-conservative forces and these measurements will be used in the drag-free control. The gravity gradient tensor is part of the differential mode (DM). The gravity gradients $V_{\beta\gamma}$ are not measured directly because the angular velocities and angular accelerations are projected onto the DM as well. The angular accelerations $\dot{\omega}_\alpha$ can be obtained by taking certain linear DM combinations, e.g.:

$$\dot{\omega}_X = \frac{a_{dYz}}{l_Y} - \frac{a_{dZy}}{l_Z} \quad (2)$$

where a_{dYz} is the z -component of \vec{a}_{dY} (see eqn. (1)), and $l_\alpha := \|\vec{r}_{1\alpha} - \vec{r}_{2\alpha}\|$. Cyclic exchange of X, Y, Z and x, y, z gives the angular accelerations around the Y and Z -axis. Integration of these accelerations gives the angular velocities ω_α , with which the gravity gradient tensor components $V_{\beta\gamma}$ can be derived in combination with the DM accelerations (Cesare, 2002):

$$V_{XX} = -2 \frac{a_{dXx}}{l_X} - \omega_Y^2 - \omega_Z^2; \quad V_{XY} = -\frac{a_{dXy}}{l_X} - \frac{a_{dYx}}{l_Y} - \omega_X \omega_Y \quad (3)$$

where again cyclic exchange gives all nine tensor elements. The accelerometer measurements are not perfect but suffer from the following errors: (1) the accelerometer is not aligned perfectly with the LORF (mis-pointing); (2) the three axes of the accelerometer are non-orthogonal (cross coupling); (3) the accelerations are measured with a scale factor; (4) the proof mass is in a motionless state somewhat away from the centre of the cage, which results in a quadratic factor; (5) the electronics readout chain introduces a bias $\vec{b}_{i\alpha}$ and noise $\underline{\vec{n}}_{i\alpha}$ (underlined values are stochastic). Because each accelerometer has two very sensitive and one less sensitive axis, three of the CM and DM accelerations will be more noisy than the other six. As a result of the accelerometer configuration, especially the diagonal gravity gradients V_{XX} , V_{YY} , V_{ZZ} have high accuracy in the measurement bandwidth (MBW, 5-100 mHz).

A model for the measured common and differential accelerations is (Willemenot, 1999; Cesare, 2002):

$$\begin{pmatrix} \vec{a}_{c\alpha} \\ \vec{a}_{d\alpha} \end{pmatrix} = \mathbf{M}_\alpha \begin{pmatrix} \vec{a}_{c\alpha} \\ \vec{a}_{d\alpha} \end{pmatrix} + \frac{\mathbf{Q}_\alpha}{2} \begin{pmatrix} (\vec{a}_{c\alpha} + \vec{a}_{d\alpha})^2 \\ (\vec{a}_{c\alpha} - \vec{a}_{d\alpha})^2 \end{pmatrix} + \begin{pmatrix} \vec{b}_{c\alpha} \\ \vec{b}_{d\alpha} \end{pmatrix} + \begin{pmatrix} \underline{\vec{n}}_{c\alpha} \\ \underline{\vec{n}}_{d\alpha} \end{pmatrix} \quad (4)$$

where the 6×6 matrices \mathbf{M}_α contain the common and differential scale factors, mis-pointings and couplings, the 6×6 matrices \mathbf{Q}_α contain the common and differential quadratic factors, and the last two terms are the common and differential bias and noise respectively. The notation of the quadratic terms means that each vector component is squared individually. Note that in this model only the measurements and the noise are stochastic quantities. Other parameters, like the elements of the matrix \mathbf{M}_α or the bias \vec{b} , may be determined by some estimation procedure from the measurements using this model. Then these estimators will be stochastic too.

It is the purpose of the *on-ground calibration* to measure the elements of \mathbf{M}_α with an error smaller than $10^{-2} - 10^{-4}$ depending on the element. The common scale factors, for example, are measured with an error of 10^{-2} . The quadratic factors of \mathbf{Q}_α are measured and physically adjusted to reduce them to zero as good as possible. In orbit, the *internal calibration* should give more accurate results than the on-ground calibration. The internal calibration is done by “shaking” the satellite using a known signal of the thrusters and extracting from the measured accelerations the part at the shaking-frequency. Again the elements of the matrices \mathbf{Q}_α are measured and physically reduced below the specified limits by proof mass position adjustment. The outcome of the internal calibration are estimates of the matrices \mathbf{M}_α :

$$\hat{\mathbf{M}}_\alpha = \mathbf{M}_\alpha + \delta\mathbf{M}_\alpha \quad (5)$$

where the differences $\delta\mathbf{M}_\alpha$ are due to measurement errors during the shaking procedure and variations in time of the accelerometer scale factors, mis-pointings and couplings.

4. External Calibration Methods

Assuming that \mathbf{Q}_α is close to zero and can be neglected, one can estimate the common and differential accelerations using $\hat{\mathbf{M}}_\alpha$ in eqn. (4). The bias is not estimated in the on-ground or internal calibration. Correct gravity gradients can therefore only be obtained using independent gravity field information. This is the role of *external calibration*. Furthermore, the mis-positioning of the accelerometers (errors in l_α) as well as the non-orthogonality of the physical gradiometer arms with respect to the {X, Y, Z}-frame do not show up in the on-ground and internal calibration. These errors, however, do propagate to the measured gravity gradients, and could perhaps be taken care of in the external calibration. An additional complicating factor is that in eqn. (2) - (3) the DM accelerations, and hence their errors, are differenced, integrated, squared, and so on. It is therefore not straightforward to see how the DM errors propagate to gravity gradient errors. Finally, external calibration offers the possibility to remove large systematic errors below the MBW that are caused by the misalignments etc. The on-ground and internal calibration are specified in such a way that the errors in the MBW are below specified levels, below the MBW the specifications are much less stringent. It is, nevertheless, desirable to produce gravity gradients as clean as possible.

Sometimes it is argued that external calibration of the gravity gradients is not necessary or not possible because for calibration *in the MBW* the existing gravity field information will most likely not be accurate enough, and for the part *below the MBW*, where the error is larger, the gradient data is complemented by the SST data which provides accurate long wavelength information. However, the gravity gradients themselves are important as a data product (to be used, e.g., directly on level 3) and it is desirable to remove systematic errors to the extent possible. Also, certain methods for gravity field determination from gravity gradients use filters to account for the coloured noise behaviour of the gradients. The filters are sensitive to the low frequency error peaks at 1, 2, ... cpr (cycles per revolution). These filters may be much more stable when the error peaks have been removed.

In principle, one may distinguish two ways of determining external calibration parameters. One way is to include calibration parameters in the gravity field inversion process, and to estimate them together with the gravity field parameters from the observations using external information to constrain the solution. This is a well-known procedure and it is based on long lasting experience. Attention should be paid to possible unwanted correlation between the calibration parameters and the gravity field parameters. The other way is to perform the external calibration in a separate pre-processing step before the gravity field inversion. Correlations as mentioned above are avoided. Care has to be taken with the selection of external data which should be of sufficient quality either in parts of the spectrum or in certain geographical regions. The latter refers to the two most promising methods for external calibration, namely by comparison of the GOCE SGG observations

with terrestrial gravity data in a regional approach or comparison with existing global gravity field models, see (Koop et al., 2001).

5. A Case Study

As an example of the estimation of calibration parameters in the pre-processing step we will conduct a GOCE-like simulation experiment. A GOCE-like orbit was generated using the EGM96 global gravity field model (Lemoine et al., 1998) truncated at degree and order 300. The orbit has an inclination of 96.6°, an average height of 250 km, and an eccentricity $e = 10^{-3}$. Gravity gradients V_{XX} , V_{YY} and V_{ZZ} in the LORF are generated along the orbit using the same model (EGM96) for a period of 30 days with a sampling interval of 5 s. Realistic errors have been generated with the SRON SGG instrument simulator, see (SID, 2000). These errors are added to the gravity gradients and the resulting gradients serve as our observations that need to be calibrated.

The gravity gradient signal PSDs exhibit strong peaks at 0, 1 and 2 cpr. These are due to the central term of the gravity field, the flattening J_2 and the coupling between the central term and the radial orbit variations. Due to the mis-pointings, cross couplings, etc. the gravity gradient errors are coupled with the signal. Additional errors are introduced in the gravity gradients due to our inability to perfectly remove the rotational term, eqn. (3). Since the rotation takes place mainly around the Y -axis, the V_{YY} errors will be smaller than the V_{XX} and V_{ZZ} errors, and will have a somewhat different characteristic. The V_{YY} errors exhibit peaks at 0-2 cpr, whereas the V_{XX} and V_{ZZ} errors show peaks at 0-4 cpr. The time series of measured gravity gradients \underline{y} (after internal calibration) may therefore be written as

$$\underline{y}(t) = \lambda[y(t) + \Delta y + y' \cdot t + \sum_{k=1}^K (a_k \cos k\omega(t) + b_k \sin k\omega(t))] + \underline{n}(t) \quad (6)$$

with $\omega = 2\pi t/T$, t the time, T the mean orbital period, $\underline{n}(t)$ the noise, and where the scale factor λ , the bias Δy , the trend y' , and the Fourier coefficients a_k, b_k are to be determined. The number of unknowns therefore is $3 + 2K$, with K in cpr. The true values y are of course not known with “infinite” precision but can be approximated using a global gravity field model. Because the long wavelength information of existing or near future models (e.g. from GRACE) is rather accurate compared to the long wavelength gravity field information content of the observed gravity gradients, it is fair to expect that external calibration using global models is able to remove systematic long wavelength errors from the gravity gradients.

Before discussing the determination of all calibration parameters in egn. (6), the scale factor determination is addressed. In (ESA, 1999) it is suggested to determine the gradiometer’s absolute scale factors using the accurate knowledge of the J_2 gravity gradients variations and the coupling between the central term and the

TABLE I

Scale factor errors $|1 - \lambda|$ estimating scale and bias.

Calibration model	V_{XX}	V_{YY}	V_{ZZ}
GM, J_2	$2 \cdot 10^{-4}$	$5 \cdot 10^{-5}$	$2 \cdot 10^{-4}$
OSU91A	$2 \cdot 10^{-7}$	$2 \cdot 10^{-5}$	$1 \cdot 10^{-5}$
“GRACE”	$7 \cdot 10^{-8}$	$9 \cdot 10^{-6}$	$4 \cdot 10^{-6}$

radial orbit variations. A direct assessment of the possibilities and limitations of that knowledge consists of the calibration of the true, errorless EGM96 gradients. In a least squares adjustment scale factors λ are determined together with a bias (the latter also compensates for the difference between the mean values of the true signal and the calibration signal). The EGM96 gradients serve as observations y , whereas the gradients generated with just the central term and J_2 serve as the calibration gradients \hat{y} . The scale factors so determined should ideally be equal to one. The scale factor errors are of the order of 10^{-4} , see Table I. The scale factor determination is repeated using OSU91A (Rapp et al., 1991) as calibration model (truncated at degree and order $L = 300$) as well as a “GRACE”-like calibration model. The latter is simply EGM96 truncated at degree and order $L = 120$, which means that the gravity field model is assumed to be errorless up to this maximum degree. The scale factor errors are 10^{-5} for V_{YY}, V_{ZZ} using OSU91A and 10^{-7} for V_{XX} . The “GRACE” model yields scale factor errors that are 2-3 times smaller than the OSU91A scale factor errors, see Table I. It is therefore better to calibrate with a full global gravity field model.

Calibration results for the simulated gravity gradients with errors are summarized in Table II. The error reduction with respect to the original, internally calibrated gravity gradients is provided. The calibration model is either OSU91A or the “GRACE” model. V_{XX} and V_{ZZ} are calibrated with 4 cpr coefficients, while V_{YY} is calibrated with 2 cpr coefficients. A first idea is to estimate the calibration parameters for the total data set, that is, for 30 days. This works reasonably well, the error reduction is significant. The error can even be further reduced if short calibration periods are considered. Calibration parameters are estimated for five periods of six days for example. The gravity gradient error is reduced by a factor 2-3. This is a consequence of the error variation in time. In the current simulation, the variation of the orbital height increases from 10 km to 30 km after 30 days. Therefore the gravity gradient signal variation increases. Because the error is coupled with the signal, the error variation increases in time, as well. Consequently, the amplitude of the Fourier coefficients of the signal and the associated error varies in time; the estimation of the Fourier coefficients in the calibration will be different for different time intervals.

The calibration period cannot be made arbitrarily small. It must be long enough for errors in the calibration model to average out. As shown in the last row of Table

TABLE II

Error reduction with respect to internal calibration:
OSU91A and (“GRACE”).

Calibration period	V_{XX}	V_{YY}	V_{ZZ}
1×30 days	0.6 (0.6)	0.8 (0.8)	0.5 (0.5)
6×5 days	0.3 (0.3)	0.5 (0.4)	0.3 (0.3)
30×1 day	0.2 (0.2)	0.7 (0.3)	0.4 (0.2)

II a calibration period of one day works well for V_{XX} whereas the error increases for V_{YY} and V_{ZZ} . The numbers in brackets give the “GRACE” model calibration results. Especially for short periods the results are much better than for OSU91A.

6. Conclusions

Calibration and validation (cal/val) are important tools to obtain high quality gravity field information from satellite gravity missions like GRACE and GOCE. Since all kinds of activities are often referred to in the context of cal/val, it is important to have a clear understanding of what we actually want to achieve with calibration and validation. With calibration we want to correct the level 1 data for systematic errors and to assess the a priori error model. Validation refers to providing a quality report of the level 2 data. It appears that cal/val for GRACE and GOCE is rather similar, although some specific aspects can be identified due to the different type of observations involved in both missions. Calibration exists of pre-launch on-ground calibration, in-flight internal calibration, and external calibration. For the latter, typically geodetic methods are to be used, such as comparisons of the observations with external gravity field knowledge. A case study of external calibration of GOCE SGG observations shows that systematic errors could be removed from the observations for the low frequencies below the MBW by comparison of the data with global gravity field models.

Acknowledgements

Most of the facts about GRACE were kindly provided to us by Mike Watkins, which is acknowledged. Details on the calibration procedures for GOCE were provided by ESA, which is acknowledged as well.

References

- Cesare, S.: 2002, 'Performance requirements and budgets for the gradiometric mission'. Issue 1 GO-TN-AI-0027, Preliminary Design Review, Alenia.
- ESA: 1999, 'Gravity Field and Steady-State Ocean Circulation Mission'. Reports for mission selection; the four candidate earth explorer core missions. ESA SP-1233(1).
- Koop, R., P. Visser, and C. Tscherning: 2001, 'Aspects of GOCE calibration'. In: *International GOCE user workshop*, Vol. WPP-188.
- Lemoine, F., S. Kenyon, J. Factor, R. Trimmer, N. Pavlis, D. Chinn, C. Cox, S. Klosko, S. Luthcke, M. Torrence, Y. Wang, R. Williamson, E. Pavlis, R. Rapp, and T. Olson: 1998, 'The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96'. TP 1998-206861, NASA Goddard Space Flight Center.
- Rapp, R., Y. Wang, and N. Pavlis: 1991, 'The Ohio State 1991 geopotential and sea surface topography harmonic coefficient models'. Report No. 410, OSU.
- SID: 2000, 'GOCE End-to-End Performance Analysis'. Final Report ESTEC Contract No. 12735/98/NL/GD, SID.
- Tapley, B. and C. Reigber: 1999, 'GRACE: a satellite-to-satellite tracking geopotential mapping mission'. In: I. Marson and H. Sünkel (eds.): *Bulletino di Geofisica Teorica ed Applicata, Sep.-Dec. 1999, Proceedings of the 2nd Joint Meeting of the International Gravity and the International Geoid Commission, Trieste 7-12 Sept. 1998, ISSN 0006-6729*, Vol. 40. p. 291.
- Velicogna, I., J. Wahr, C. Milly, M. Shaltout, and A. ArRajehi: 2000, 'Validation of GRACE Satellite Geoid Measurements'. Poster presented at AGU Fall Meeting, San Francisco.
- Willemenot, E.: 1999, 'Capacitive gradiometer: Calibration and verification approach'. Concluding Technical Report RTS 19/3825 DMPH/Y, Onera.

V: SEA LEVEL

BENEFITS OF GRACE AND GOCE TO SEA LEVEL STUDIES

PHILIP L. WOODWORTH (plw@pol.ac.uk)

Proudman Oceanographic Laboratory, Bidston Observatory, Birkenhead CH43 7RA, U.K.

JONATHAN M. GREGORY (jonathan.gregory@metoffice.com)

Hadley Centre for Climate Prediction and Research, Bracknell RG12 2SY, U.K.

Received: 1 May 2002; Accepted in final form: 1 August 2002

Abstract. The recently published Third Assessment Reports of the Intergovernmental Panel on Climate Change have underlined the scientific interest in, and practical importance of, past and potential future sea level changes. Space gravity missions will provide major benefits to the understanding of the past, and, thereby, in the prediction of future, sea level changes in many ways. The proposal for the GOCE mission described well the improvements to be expected from improved gravity field and geoid models in oceanography (for example, in the measurement of the time-averaged, or ‘steady state’, ocean surface circulation and better estimation of ocean transports), in geophysics (in the improvement of geodynamic models for vertical land movements), in geodesy (in positioning of tide gauge data into the same reference frame as altimeter data, and in improvement of altimeter satellite orbits), and possibly in glaciology (in improved knowledge of bedrock topography and ice sheet mass fluxes). GRACE will make many important steps towards these ‘steady state’ aims. However, its main purpose is the provision of oceanographic (and hydrological and meteorological) temporally-varying gravity information, and should in effect function as a global ‘bottom pressure recorder’, providing further insight into the 3-D temporal variation of the ocean circulation, and of the global water budget in general. This paper summaries several of these issues, pointing the way towards improved accuracy of prediction of future sea level change.

1. Introduction

The scientific case for the Gravity Field and Steady State Ocean Circulation Explorer (GOCE) mission at the European Space Agency ‘Granada-2’ conference was constructed to a great extent around the many benefits of better knowledge of the Earth’s gravity field and geoid to the measurement and understanding of sea level change (Balmino et al., 1999). In particular, it was recognised that a better geoid would result in major gains in understanding the ocean circulation and solid Earth and glaciological processes which contribute to sea level change. Improved understanding of the processes then might lead to improvements in our ability to predict future changes.

The case for a temporal gravity mission such as the Gravity Recovery and Climate Experiment (GRACE) was also constructed partly around the sea level change issue, by providing a better understanding of the global hydrological cycle, through the monitoring of changes in gravity associated with ground water and ice



caps, and of the ocean thermohaline circulation, through the monitoring of spatial variations in ocean water mass (or bottom pressure) (NRC, 1997; GRACE, 1998).

This paper provides an overview of our knowledge of global sea level change during the 20th century, including a review of the various terms which contribute to the change. It draws extensively on the conclusions of the sea level group of the recent Intergovernmental Panel on Climate Change Third Assessment Report (IPCC TAR) (Church et al., 2001). Other papers in the present volume will expand further on the potential of space gravity to provide information on some of the terms.

2. The Past 100 Years

Most of our knowledge of global sea level change during the 20th century comes from the tide gauge data set of the Permanent Service for Mean Sea Level (PSMSL) (Woodworth, 1991). The balance of evidence from this data set, derived from the work of a number of authors, is that global sea level increased at an average rate of 1–2 mm/year during the past 100 years, with some evidence for a small acceleration in the rate of sea level rise between the 19th and 20th centuries. The magnitude of this recent change can be compared to the 120 m or so of sea level rise which occurred since the last glacial maximum, and is comparable to the amplitude of fluctuations on timescales of a few hundred years or longer which could have taken place during the past 6000 years.

A particular problem with tide gauge data, as with geological, archaeological or any other sea level data obtained from the Earth's surface, is that measured rates are 'relative' ones. That is to say, they are measures of sea level change relative to the level of the land upon which the gauge is located, and where 'land level' is defined by a set of benchmarks near to the gauge. The tide gauge community is investing heavily in new geodetic techniques by means of which vertical land movements (VLMs) at the benchmarks will be monitored, thereby decoupling 'real' sea and land signals in the relative tide gauge records. These techniques are primarily the Global Positioning System (GPS), Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), and Absolute Gravity. However, this is still a developing field and, while the techniques have clearly shown their potential (e.g. see Bingley et al., 2001; Cazenave et al., 1999; Williams et al., 2001), results on VLMs at gauges obtained this way have not so far been applied to a study of global sea level change. Consequently, the authors of papers considered by the IPCC TAR had once again applied land movement corrections using the same techniques which have been used for some years.

Two VLM-correction techniques are often used. One concerns the application of geodynamic numerical models of Glacial Isostatic Adjustment (GIA). GIA is the only geological process leading to VLM for which we have global models which can be applied to tide gauge data, although there are significant differences

between the models of different authors in ice history, viscosity parameterisation etc. In addition, there are obviously other geological processes which can lead to VLM in certain regions (e.g. active tectonics, Di Donato et al., 1999). The second concerns the use of geological (i.e. long term, typically since 6K BP) sea level data obtained from sites near to the gauges (e.g. Woodworth et al., 1999). By subtracting the long term rate of change of sea level, from whatever geological cause, from the tide gauge rate, one obtains estimates of sea level change due to present-day processes (e.g. climate change). The main difficulty with this method is that sufficiently reliable and copious geological data are not available worldwide.

Peltier (2001) and Church et al. (2001) contain detailed discussions of the relative merits of each of the VLM-correction techniques. However, in the context of the present paper, it is already clear from the GRACE and GOCE proposals that improvements in the models of GIA and active tectonics from missions such as these would be a major first benefit to our analysis of tide gauge data from the past 100 years, with the potential for removing a large part of the 1 mm/year range of uncertainty quoted by Church et al. (2001).

If one takes the 1–2 mm/year rate of global sea level change for the past 100 years at face value, then the next question concerns which processes might have contributed to it. Figure 1 lists possible processes together with minimum, mean and maximum estimates of their contribution. For example, by considering an ensemble of hydrographic measurements for the past 100 years together with a set of results from atmosphere-ocean General Circulation Models (GCMs) forced by observed global air temperature changes, one concludes that thermal expansion could have contributed as little as 3 cm and as much as 7 cm to global sea level change. Other processes can be seen to contribute at a similar or lower rate (and for a full discussion, see Church et al., 2001), except for terrestrial water storage (hydrology), for which the range of uncertainty is very large.

The sum of the various processes amounts to a hindcast of global sea level change for the past 100 years of +7 cm, within a range of –8 to +22 cm, which could be consistent with the observed amount, if a little on the low side.

Let us speculate how the IPCC fourth and fifth reviews, which one expects might be held around 2006 and 2011, might benefit from knowledge from GRACE and GOCE and improve on the TAR, first in refining the estimates of the change over the past 100 years, and second in understanding them. Table I shows that the first benefit will stem from the improvements in GIA models discussed above. Even though one expects that by 2011 many gauges will have acquired GPS land movement time series more than one decade long, not all gauges by any means will have done so. In particular, it cannot be expected that the same density of GPS measurements will be undertaken in Africa, South America etc. as in Europe and USA. Consequently, GIA modelling will continue to be the only source of VLM information for many sites.

A second benefit might be expected from connections of short historical tide gauge records obtained some distance (perhaps few 10s km) apart. Many long tide

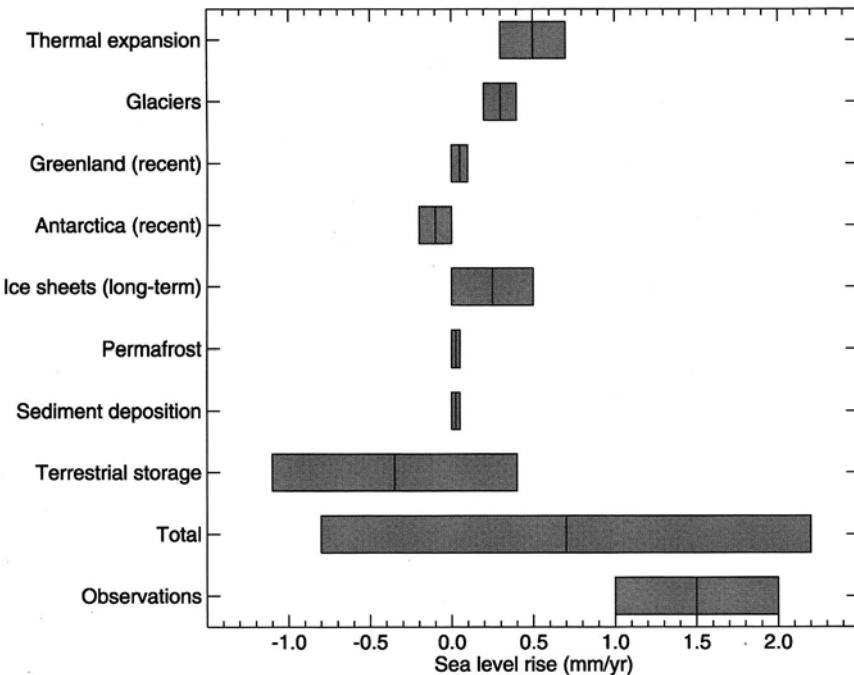


Figure 1. Ranges of uncertainty for the average rate of sea level rise from 1910–1990 and the estimated contributions from different processes. From Church et al. (2001).

gauge records (e.g. Brest, San Francisco) are in fact composites, derived from gauges which were installed in different parts of a port at different times and which were relocated as harbour developments took place. The construction of composites has been possible because their separate benchmarks were preserved and connected by conventional levelling. However, there has been a practical limit (few km typically) for which such port levelling has usually been performed. Consequently, in some countries there exist records which one could combine into composites if one could reliably undertake ‘GPS-geoid levelling’ using improved geoid models. Some benefits might also accrue from improvements to the orbits of older altimeter satellites (e.g. Geosat, ERS-1/2). The benefits would necessarily be limited, as the main difficulty in improving the accuracy of the older satellite orbits is concerned with the quantity and quality of tracking data available, rather than gravity field errors. In addition, their data apply only to the mid-1980s onwards. Nevertheless, some improvements from improved orbits, combined by more detailed *in situ* calibration studies using gauges, can be expected.

GRACE and GOCE might be more beneficial in improving our understanding of the processes which have contributed to sea level change during the past 100 years. Table I lists the processes from Figure 1, divided into those which might possibly accrue and those which are unlikely to benefit. For example, one can imagine that our improved understanding of the thermohaline circulation will re-

TABLE I

Sea Level Observations and Processes of the Past 100 Years: Benefits from Space Gravity

Potential Improvements in Observational Data
GIA model improvements at gauge sites
GPS-geoid ‘levelling’ connections between short tide gauge records
Improved satellite orbits for historical altimetry
Potential Improvements in Understanding Processes
<i>Possible:</i>
Thermal expansion
Greenland
Antarctica
<i>Unlikely:</i>
Glaciers
Permafrost
Sediment deposition
Terrestrial water storage

fine GCM estimates of thermal expansion for the past 100 years. In addition, the processes involved in polar ice sheet mass balance are primarily long term ones. Consequently, the understanding we obtain from GRACE could feasibly be applied to the past 100 years. In the case of terrestrial water storage, for which GRACE is particularly suited, one could imagine increased understanding in, for example, seasonal cycles of water storage, or changes in hydrology connected to El Niño. However, it is hard to see how GRACE could contribute to reduction of the large uncertainties associated with dam construction or farming practices over the past 100 years, which result in the large range of uncertainty for terrestrial water storage shown in Figure 1.

As a digression, we should point to three important papers published since the IPCC TAR which cast further light on the different processes responsible for past sea level change. Tamisiea et al. (2001) discuss a ‘fingerprint’ method, wherein spatial variations in sea level trends are used to estimate global sea level change arising from either mountain glaciers, Greenland and ‘Antarctic plus other’. (This paper also reminds us that the often-misused word ‘eustatic’ means change in ocean volume, not a uniform sea level change.) Their results suggest a higher rate of Greenland melting than that suggested in the Church et al. (2001) review, and might offer an explanation for the lower rates of sea level rise found in Europe than

elsewhere (Woodworth et al., 1999). Plag and Jüttner (2001) adopted a similar approach, finding the magnitude and sign of estimated Greenland and Antarctic contributions to be sensitive to the latitudinal selection of tide gauge trends employed in the analysis. Cabanes et al. (2001) made use of the hydrographic data set of Levitus spanning 1955–95 to estimate changes in global sea level due to thermal expansion, with a conclusion that the 1–2 mm/year rate of Church et al. may have been over-estimated. However, this conclusion depends critically on the validity of using the sparse hydrographic data set in this way (particularly given only temperature and not salinity information), and of implying change on century time scales from a shorter record (Douglas and Peltier, 2002).

3. The Next 100 Years

Discussion of changes in global sea level during the next 100 years is almost entirely a GCM-based one. Church et al. (2001) provide estimates of change under a range of emission scenarios (depending on world population, economic activity etc.), climate sensitivities (the amount of global warming for a given greenhouse gas forcing), and GCM formulation. The overall prediction is for a rise of between 9 and 88 cm between 1990–2100, with a central value of 48 cm. This represents a rate of approximately 2–4 times that of the past 100 years which could result in significant impacts in many low-lying coastal areas.

Let us again consider first the monitoring system over the next 100 years. Of course, gauges will still be required for the study of long term trends and for altimeter calibration, and gauges will continue to be essential at the coast, which is where people live. However, the community will undoubtedly see altimetry replacing tide gauges as the main global, or deep-sea, monitoring system (Mitchum et al., 2001).

Table II shows that the gravity missions will help the altimeter missions in two main ways. Geographically-correlated altimeter orbit errors should almost completely vanish (Table 3.1 of Balmino et al., 1999). In addition, uncertainties in the small GIA corrections which are necessary to make to global sea level measurements should be removed. The corrections are of order 0.3 mm/year and reflect the secular change in the global-average geoid (Peltier, 2001).

However, if one compares the benefits from these terms (perhaps a few cm at most in each case) to the anticipated increase in global sea level in the next 100 years (48 cm), it is clear that the community already has a good monitoring system, based primarily upon TOPEX/Poseidon class altimetry, and that any benefits will be limited ones. If the predictions are correct, we simply have to wait for the global sea level to rise.

Table II lists the various processes responsible for the sea level change, with their potential for better understanding, once again divided into ‘possible’ and ‘unlikely’. Terrestrial water storage, a particular scientific objective of GRACE and future temporal gravity missions, is now definitely a ‘possible’. Glaciers have been

TABLE II

Sea Level Observations and Processes of the Next 100 Years: Benefits from Space Gravity

Potential Improvements in Observational Data
Gravity field-related orbit errors of altimeter satellites completely removed
GIA model improvements for global sea level correction
Potential Improvements in Understanding Processes
<i>Possible:</i>
Thermal expansion (and its spatial variation)
Greenland
Antarctica
Terrestrial water storage
<i>Unlikely:</i>
Glaciers etc.

shown as ‘unlikely’, which is the case from a space gravity perspective. However, one anticipates that developments in various kinds of altimetry will eventually provide data on the mass balance of the world’s glaciers. As a consequence, it should be possible to monitor, if not always to understand perfectly, most of the important processes of Figure 1 in the future.

Two special points must be made about thermal expansion. The first can be made from Table III (adapted from Table 11.14 of Church et al., 2001) which shows results from one GCM exercise in which the predicted sea level rise 1990–2100 was obtained using a common emission scenario (IS92a, including the direct effect of sulphate aerosols) using the models of different groups. It can be seen that, in most model runs, thermal expansion is clearly the most important process for the next 100 years, and is consequently one which needs to be fully understood.

The second point about thermal expansion is that it will not be spatially uniform, but will vary regionally as the ocean circulation attempts to adjust to the changing fluxes (Church et al., 2001; Gregory et al., 2001). On one hand this is a problem; it means that the impacts of sea level change could be more serious along some coastlines than others. On the other hand, the 3-D changes reflected in spatially-varying sea level and bottom pressure (measured by altimetry and space gravity together with *in situ* systems such as ARGO) will provide a signal one can use to learn how the ocean works on decadal and longer time scales.

TABLE III

Sea-level rise 1990–2100 due to climate change derived from GCM experiments following the IS92a scenario, including the direct effect of sulphate aerosols. Expansion, Min Sum and Max Sum indicate the thermal expansion and minimum and maximum overall sea level change respectively. See Church et al. (2001) for estimates of the other, smaller terms which contribute to Min Sum and Max Sum.

Experiment	Expansion, m	Min Sum, m	Max Sum, m
CGCM1 GS	0.43	0.53	0.70
CSIRO Mk2 GS	0.33	0.36	0.53
ECHAM4/OPYC3 GS	0.30	0.26	0.41
GFDL R15a GS	0.38	0.43	0.60
HadCM2 GS	0.23	0.27	0.42
HadCM3 GSIO	0.24	0.24	0.40
MRI2 GS	0.11	0.14	0.27
DOE PCM GS	0.20	0.17	0.32
Range		0.14	0.70

4. Towards Better GCM Predictions

The problem with GCM predictions of global sea level change such as those of Church et al. (2001) is that they are at the end of a very long line of arguments and assumptions concerning emission scenarios and parameterisations of chemical and physical processes in the environment which lead to a climate change. Ganachaud et al. (1997), Le Grand and Minster (1999), Le Grand (2001) and Schröter et al. (2002) have simulated the benefits of an improved geoid model from a mission such as GOCE to the determination of ocean fluxes. However, we need to know how that increased ocean understanding feeds through to the GCMs.

Two examples can be given of clues on how the GCM community might make some progress. The first example, taken from Raper et al. (2002), is given by Figure 2 which shows the ocean heat uptake efficiency for several models (almost the same set as in Table III) using the same CMIP2 emission scenario. In this scenario, CO₂ increases by 1 percent per year compounded for 80 years, with 2×CO₂ reached after 70 years. The efficiency is defined as the average net heat flux change into the ocean for years 60–80 divided by the average surface air temperature change (with respect to a zero CO₂ increase control run). The division factors out the different climate sensitivities of the models, which is responsible for a lot of the spread in thermal expansion between models. However, quite a considerable spread remains, implying that model physics differs to the extent that heat is transported across and into the ocean in different ways.

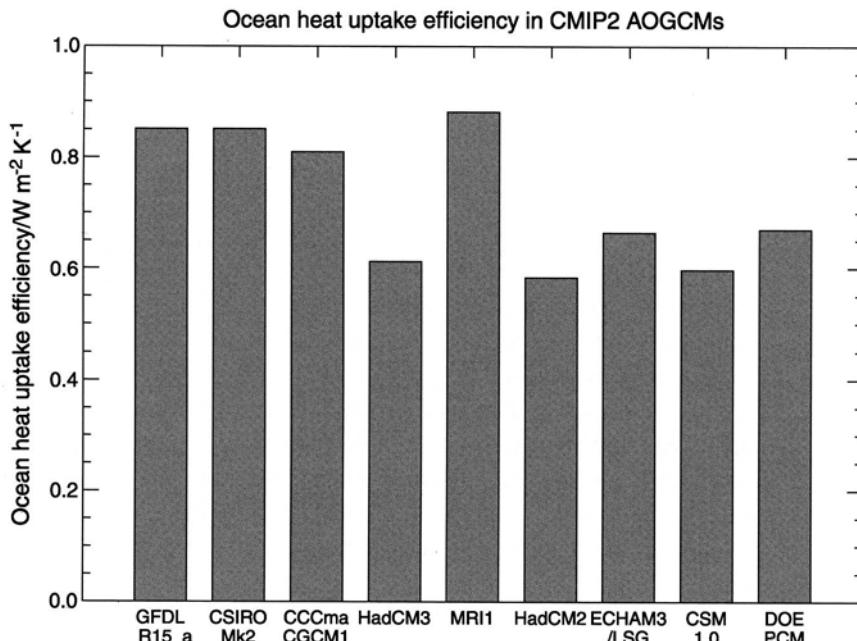


Figure 2. Ocean heat uptake efficiency for several GCMs using the same emission scenario (CMIP2).

A second example is given by Figure 11.15 of Church et al. (2001) which shows different scenario and model predictions for global sea level change for the next 3000 years. (The models in this case are of ‘intermediate complexity’ and not full GCMs.) The first thing one spots is that in some runs a rise of several metres over several 100 years is predicted, which would have devastating effects on many coastal areas of the world. However, one also sees a big spread between models. Again, some of this spread is because of different climate sensitivities and hence different surface warmings, but some is due to different amounts of heat taken up by the deep ocean. For example, the different versions of the University of Victoria, Canada (UVic) model have the same warming in a given scenario, but very different thermal expansion.

These two examples are encouraging in a way, in that, if there are big differences between models, then sea surface topography (altimetric mean sea surface minus geoid) alone might be informative in selecting between them. However, the more we know of how the ocean itself works, the better we can construct constraints on volumetric, heat and freshwater fluxes which in turn can be used to constrain the choice of parameters such as those controlling ocean diffusivities and atmosphere-ocean exchanges within the GCMs. As the ocean components of GCMs used for climate studies have rather coarse resolution (typically 1–5 degrees), they rely heavily upon large-scale parameterizations of small-scale processes, and the schemes used for this necessarily contain values which are difficult to relate directly to observations, but which are adjusted within plausible limits to

obtain a match to reality. These iterations of understanding will, of course, have benefits to climate prediction beyond the issue of sea level change alone.

5. Conclusions

This paper has attempted a brief review of findings on the sea level changes of the past 100 years and on the predictions for the next 100 years. It is clear from previously published work, and from presentations in this volume, that space gravity can help in many ways towards an understanding of the ‘sea level problem’. In particular, both the GRACE and GOCE types of space gravity have the potential to provide increased understanding of ocean circulation, and of improved prediction of ocean thermal expansion which will be a major component of any future global sea level change. This has been recognised in recent important documents such as the ‘Reducing the Uncertainties’ section of Church et al. (2001). Exactly by how much the uncertainties can be reduced remains the topic of further research.

Acknowledgements

JMG’s work was supported by the UK Department of the Environment, Food and Rural Affairs under contract PECD 7/12/37 and by the UK Government Meteorological Research Contract.

References

- Balmino, G., R. Rummel, P. Visser, and P. Woodworth: 1999, ‘Gravity Field and Steady-State Ocean Circulation Mission. Reports for assessment: the four candidate Earth Explorer Core Missions’, European Space Agency Report SP-1233(1), 217pp.
- Bingley, R.M., A.H. Dodson, N.T. Penna, F.N. Teferle, and T.F. Baker: 2001, ‘Monitoring the vertical land movement component of changes in mean sea level using GPS: results from tide gauges in the UK’, *Journal of Geospatial Engineering*, **3**, 9–20.
- Cabanes, C., A. Cazenave, and C. Le Provost: 2001, ‘Sea level rise during past 40 years determined from satellite and in situ observations’, *Science*, **294**, 840–842.
- Cazenave, A., L. Soudarin, J-F. Cretaux, and C. Le Provost: 1999, ‘Sea level changes from Topex-Poseidon altimetry and tide gauges, and vertical crustal motions from DORIS’, *Geophysical Research Letters*, **26**, 2077–2080.
- Church, J.A., J.M. Gregory, P. Huybrechts, M. Kuhn, K. Lambeck, M.T. Nhuan, D. Qin, and P.L. Woodworth: 2001, ‘Changes in sea level’. In: J.T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P.J. Van Der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.): *Climate Change 2001: The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge: Cambridge University Press, pp. 639–693.
- Di Donato, G., A.M. Negredo, R. Sabadini, and L.L.A. Vermeersen: 1999, ‘Multiple processes causing sea-level rise in the central Mediterranean’, *Geophysical Research Letters*, **26**, 1769–1772.

- Douglas, B.C. and W.R. Peltier: 2002, 'The puzzle of global sea-level rise', *Physics Today*, **55**(3), 35–40.
- Ganachaud, A., C. Wunsch, M.-C. Kim, and B. Tapley: 1997, 'Combination of TOPEX/Poseidon data with a hydrographic inversion for determination of the oceanic general circulation', *Geophysical Journal International*, **128**, 708–722.
- GRACE, 1998. 'Gravity Recovery and Climate Experiment science and mission requirements document', Revision A, JPLD-15928, NASA's Earth System Science Pathfinder Program.
- Gregory, J.M., J.A. Church, G.J. Boer, K.W. Dixon, G.M. Flato, D.R. Jackett, J.A. Lowe, S.P. O'Farrell, E. Roeckner, G.L. Russell, R.J. Stouffer, and M. Winton: 2001, 'Comparison of results from several AOGCMs for global and regional sea-level change 1990–2100', *Climate Dynamics*, **18**, 225–240.
- Le Grand, P. and J.-F. Minster: 1999, 'Impact of the GOCE gravity mission on ocean circulation estimates', *Geophysical Research Letters*, **26**, 1881–1884.
- Le Grand, P.: 2001, 'Impact of the gravity field and steady-state ocean circulation explorer (GOCE) mission on ocean circulation estimates. Volume fluxes in a climatological inverse model of the Atlantic', *Journal of Geophysical Research*, **106**, 19597–19610.
- Mitchum, G.T., R. Cheney, L.-L. Fu, C. Le Provost, Y. Menard, and P. Woodworth: 2001, 'The future of sea surface height observations'. In: C.J. Koblinsky and N.R. Smith (eds.): *Oceans in the 21st Century*, Melbourne, Australia: Bureau of Meteorology, pp. 120–136.
- NRC, 1997, 'Satellite gravity and geosphere', Washington, D.C.: National Academy Press.
- Peltier, W.R.: 2001, 'Global glacial isostatic adjustment'. In: B.C. Douglas, M.S. Kearney, and S.P. Leatherman (eds.): *Sea level rise: history and consequences*, San Diego: Academic Press, pp. 65–95.
- Plag, H.-P. and H.-U. Jüttner: 2001, 'Inversion of global tide gauge data for present-day ice load changes'. In: T. Yamanouchi (ed.): *Proceedings of the Second International Symposium on Environmental Research in the Arctic and Fifth Ny-Alesund Scientific Seminar*, Memoirs of the National Institute of Polar Research, Special Issue No. 54, 301–317.
- Raper, S.C.B., J.M. Gregory, and R.J. Stouffer: 2002, 'The role of climate sensitivity and ocean heat uptake on AOGCM transient temperature response', *Journal of Climate*, **15**, 124–130.
- Schröter, J., M. Losch, and B. Sloyan: 2002, 'Impact of the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) mission on ocean circulation estimates 2. Volume and heat fluxes across hydrographic sections of unequally spaced stations', *Journal of Geophysical Research*, **107**, 10.1029/2000JC000647.
- Tamisiea, M.E., J.X. Mitrovica, G.A. Milne, and J.L. Davis: 2001, 'Global geoid and sea level changes due to present-day ice mass fluctuations', *Journal of Geophysical Research*, **106**, 30849–30863.
- Williams, S.D.P., T.F. Baker, and G. Jeffries: 2001, 'Absolute gravity measurements at UK tide gauges', *Geophysical Research Letters*, **28**, 2317–2320.
- Woodworth, P.L.: 1991, 'The Permanent Service for Mean Sea Level and the Global Sea Level Observing System', *Journal of Coastal Research*, **7**, 699–710.
- Woodworth, P.L., M.N. Tsimplis, R.A. Flather, and I. Shennan: 1999, 'A review of the trends observed in British Isles mean sea level data measured by tide gauges', *Geophysical Journal International*, **136**, 651–670.

Address for Offprints: P.L. Woodworth, Proudman Oceanographic Laboratory, Bidston Observatory, Birkenhead, Merseyside CH43 7RA, U.K., plw@pol.ac.uk

WHAT MIGHT GRACE CONTRIBUTE TO STUDIES OF POST GLACIAL REBOUND?

JOHN WAHR and ISABELLA VELICOGNA

*Department of Physics and Cooperative Institute for Research in Environmental Sciences
University of Colorado, Boulder, CO*

Received: 16 May 2002; Accepted in final form: 22 July 2002

Abstract. The NASA/DLR satellite gravity mission GRACE, launched in March, 2002, will map the Earth's gravity field at scales of a few hundred km and greater, every 30 days for five years. These data can be used to solve for time-variations in the gravity field with unprecedented accuracy and resolution. One of the many scientific problems that can be addressed with these time-variable gravity estimates, is post glacial rebound (PGR): the viscous adjustment of the solid Earth in response to the deglaciation of the Earth's surface following the last ice age.

In this paper we examine the expected sensitivity of the GRACE measurements to the PGR signal, and explore the accuracy with which the PGR signal can be separated from other secular gravity signals. We do this by constructing synthetic GRACE data that include contributions from a PGR model as well as from a number of other geophysical processes, and then looking to see how well the PGR model can be recovered from those synthetic data. We conclude that the availability of GRACE data should result in improved estimates of the Earth's viscosity profile.

1. Introduction

GRACE, jointly sponsored by NASA and the Deutsches Zentrum für Luft- und Raumfahrt (DLR), was launched in March, 2002. It will map the Earth's gravity field with unprecedented accuracy and resolution every 30 days during its 5-year lifetime. This will permit monthly variations in gravity to be determined down to scales of a few hundred kilometers and larger. These gravity variations can be used to study a variety of processes involving redistribution of mass within the Earth or on its surface. The expected performance of GRACE and various possible applications are described by Dickey et al. (1997) and Wahr et al. (1998).

Among these applications is the use of the GRACE secular gravity signal to constrain models of post glacial rebound (PGR): the viscous adjustment of the solid Earth in response to the removal of the ice loads following the last ice age. PGR studies are useful from a solid Earth perspective, because they provide information about the Earth's viscosity profile. The PGR process is also an error source when interpreting various types of observations relevant to global sea level change, including altimeter estimates of ice sheet thickness variability and tide gauge estimates of the sea level change itself.



This paper describes a preliminary look at GRACE's ability to recover the PGR signal. A more detailed description can be found in Velicogna and Wahr (2002a). We will focus primarily on the time-variable gravity signal over northern Canada. It will probably be harder for GRACE to constrain the PGR signal over Scandinavia, due to the smaller amplitudes and shorter spatial scales that characterize the signal there. The PGR gravity signals over Greenland and Antarctica are likely to be severely contaminated by the gravity signals from the present-day mass imbalance of the Greenland and Antarctic ice sheets. Methods of combining GRACE and ice sheet altimeter data to separate the PGR and ice sheet signals in those regions are described by Wahr et al. (2000), Velicogna and Wahr (2002b), and Wu et al. (2002).

2. The PGR Signal

A PGR model requires knowledge of the Earth's viscosity profile and of the ice deglaciation history. These are determined by comparing PGR model output and observations. Specific types of observations tend to be more sensitive to particular parameters. For example, the viscosity of the Earth's upper mantle can be especially well constrained using geological observations of past changes in relative sea level, and geodetic observations of present-day crustal uplift. Observations more sensitive to lower mantle viscosity include the present-day free-air gravity and geoid anomalies over Canada and Scandinavia, and secular changes in the Earth's rotation and in the earth's gravity field as determined from satellites.

Past interpretations of the observations most sensitive to lower mantle viscosity have been somewhat ambiguous. The present-day gravity and geoid anomalies may have sizable tectonic contributions, unrelated to PGR. The Earth rotation and time-variable gravity observations could well have significant contributions from other quasi-secular processes, including present-day changes in polar ice.

In this paper, we investigate whether time-variable gravity from GRACE will have the resolution and accuracy needed to separate the PGR effects from the contaminating effects of other secular processes. We do this by constructing simulated GRACE data using output from a PGR model, and then trying able to use those GRACE data to recover the input model used in the simulation.

For our estimate of the deglaciation history we use the global ICE-3G Pleistocene ice model of Tushingham and Peltier (1991), with an additional 90 kyr linear glaciation phase added at the beginning. Our viscosity profile consists of three uniform viscosity layers: the uppermost mantle (between the base of the lithosphere and 400 km depth), the transition zone (between 400 km and 670 km depth), and the lower mantle (between the core-mantle boundary and 670 km depth). The overlying lithosphere and underlying fluid core are assumed to be inviscid. We will simulate results for many viscosity values and lithospheric thicknesses. For our "default" model, we assume a lithospheric thickness of 100 km, uppermost mantle and transition zone viscosities of 1.0×10^{21} Pa s, and a lower mantle viscosity of

TABLE I
Elastic Parameters Used to Compute All PGR Contributions

Layer	Outer radius (R , km)	Density (ρ , kg/m ³)	Shear wave speed (v_s , km/s)
Core	3480	10,925	0.0
Lower Mantle	5701	4,970	6.6
Transition Zone	5971	3,850	5.25
Uppermost Mantle	6271	3,070	4.33
Lithosphere	6371	3,070	4.33

1.0×10^{22} Pa s. For the elastic structure, we assume the five spherical layers are incompressible and homogeneous. Their densities and shear wave velocities are chosen to be reasonably consistent with the structural model PREM (Preliminary Reference Earth Model) of Dziewonski and Anderson (1981), and are given in Table I.

3. Time Variable Gravity from GRACE

It is usual to expand the geoid height, N , as a sum of associated normalized Legendre functions, \tilde{P}_{lm} , in the form (see, *e.g.*, Chao and Gross, 1987):

$$N(\theta, \phi) = a \sum_{l=0}^{\infty} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) [C_{lm} \cos(m\phi) + S_{lm} \sin(m\phi)], \quad (1)$$

where θ and ϕ are co-latitude and eastward longitude, and the C_{lm} 's and S_{lm} 's are dimensionless Stokes' coefficients. GRACE measurements will be used to determine the C_{lm} 's and S_{lm} 's up to degree and order (*i.e.* l and m) = 100 every 30 days. For each \tilde{P}_{lm} term in this expansion, the horizontal scale (half-wavelength) is approximately $20,000/l$ km.

PGR causes secular variations in the C_{lm} 's and S_{lm} 's. But so do other processes. Present-day changes in the Greenland and Antarctic ice sheets pose the most serious problems for using GRACE to learn about PGR. They are likely to be the largest non-PGR sources of secular gravity, and they are located at latitudes similar to PGR latitudes. This has, historically, made it difficult to separate these signals in the time-variable satellite gravity field.

Prior to the launch of GRACE, all satellite estimates of time-variable gravity came from satellite laser ranging (SLR). The SLR secular terms consist of a few zonal coefficients: *i.e.* C_{l0} 's for $l = 2$ up to maybe $l = 6$ (see, for example, Cheng et al., 1997). This is not enough coefficients to allow for easy separation of

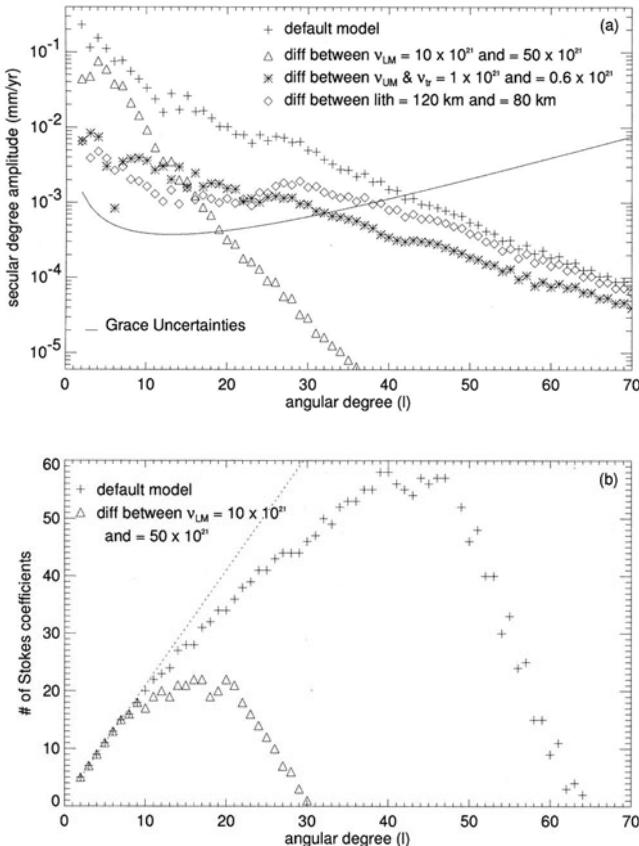


Figure 1. (a) Predictions of degree amplitudes for the secular change in the geoid, as a function of angular degree l . Shown are predictions for the default PGR model (plus signs), and for the differences between the default PGR model and three other plausible PGR models (v_{UM} , v_{LM} , and v_{Tr} are the viscosities, in Pa s, of the uppermost mantle, lower mantle, and transition zone; and lith is the lithospheric thickness). Also shown (solid line) are the degree amplitudes of the expected secular GRACE measurement errors. (b) The number of Stokes' coefficients at a given value of l where the amplitude of the predicted PGR signal is larger than the expected secular GRACE measurement error. Shown are the results for the default PGR model, and for the difference between the PGR signals for two plausible lower mantle viscosity values. The dotted line shows the total number of Stoke's coefficients as a function of l . All together, there are a total of 2193 Stokes' coefficients for the default PGR signal, and 424 coefficients for the difference between the two PGR signals, that rise above the expected GRACE errors.

PGR from the competing secular signals. For example, the difference in longitude between Hudson Bay (the center of the largest Pleistocene ice sheet) and Greenland is about 40° . Separation of these two signals would thus require angular orders up to about $m = 360^\circ/40^\circ = 9$. Zonal coefficients ($m = 0$) alone, do not permit this separation.

GRACE, though, will have considerably higher spatial resolution and accuracy. Figure 3a compares the secular degree amplitudes, defined as

$$N_l = a \sqrt{\sum_{m=0}^l (C_{lm}^2 + S_{lm}^2)} , \quad (2)$$

of the default PGR geoid signal (the plus signs), with the expected degree amplitudes of the secular GRACE measurement errors as estimated by B. Thomas and M. Watkins [personal communication]. The results in Figure 3a show the PGR signal should be larger than the GRACE measurement errors for degrees of about 40 and smaller. Since there are $2l + 1$ values of m for every l , this suggests there should be about 40^2 Stokes' coefficients where the PGR signal is larger than the secular GRACE errors. In fact, there are 2193 such coefficients, shown in Figure 3b (the plus signs) as a function of l .

Given this high sensitivity of GRACE, it is natural to also consider GRACE's sensitivity to the effects of different viscosity profiles. The other symbols in Figure 3a show the degree amplitudes of the differences between pairs of PGR models. For each of the three PGR pairs considered in this figure, one model is the default model, and the other is obtained from the default model by changing either the lower mantle viscosity, the uppermost mantle and transition zone viscosity, or the lithospheric thickness. Note that for low angular degrees, where GRACE is most accurate, the PGR results are most sensitive to the lower mantle viscosity. The triangles in Figure 3b show there are a large number of Stokes' coefficients (a total of 424) where the difference between PGR models for two different but plausible lower mantle viscosities rise above the expected GRACE errors.

4. GRACE Simulations

The results shown in Figure 3 are encouraging, because they imply that the GRACE measurement accuracy should easily be good enough for GRACE to be sensitive to the PGR signal. But those results say nothing about the problem of separating the PGR signal from other gravity signals that appear to be secular over the five year lifetime of GRACE. To address that issue, we construct a suite of plausible, simulated GRACE data sets that include time-variable gravity signals from a number of sources, including PGR. We look to see how well we can recover the PGR signal from each of those simulated data sets.

Each simulated GRACE data set is in the form of five years of monthly-averaged Stokes' coefficients. In those coefficients we include a realization of the GRACE measurement errors estimated by Thomas and Watkins, as well as the predicted gravity signals from PGR (using our default model); from spatially-uniform, secular ice mass changes over Antarctica and Greenland (estimated assuming sea level rise contributions from Antarctica and Greenland that are within the range of $1 \pm$

1 mm/yr and 0.1 ± 0.15 mm/yr, respectively – numbers consistent with the mass balance estimates summarized by Church et al. (2001); from a 1.0 mm/yr non-steric global sea level rise; from month-to-month variations in global, continental water storage (provided by C. Milly and K. Dunne, personal communication); from month-to-month variations in sea floor pressure caused by changes in oceanic circulation (estimated by M. Molenaar and F. Bryan, personal communication) from the POP ocean general circulation model (Dukowicz and Smith, 1994); and from errors in the GRACE atmospheric pressure corrections over land, estimated by taking the differences between ECMWF (European Centre for Medium-Range Weather Forecasts, 1995) and NCEP (Kalnay et al., 1996) pressure fields, divided by $\sqrt{2}$.

Once we have constructed the simulated GRACE data, we use two methods to get an idea of how well the default PGR model could be recovered from the data. Both methods fall short of the full inversion that people will undoubtedly use with the real GRACE data. But they establish some initial expectations for how well the PGR signal can be separated from other sources of secular gravity.

4.1. GRACE RECOVERY OF PGR: METHOD 1

As a first attempt to look at the effects of contamination, we simply least squares fit the default PGR signal to the simulated GRACE data. In the absence of contamination, and if there were no GRACE errors, the fit parameter would equal 1.

Figure 4.1a shows the secular geoid change inferred from one of the simulated GRACE data sets; where the sea level rise contributions from Antarctica and Greenland are equal to 1 mm/yr and 0.1 mm/yr, respectively. Figure 4.1a thus represents a secular geoid signal that might be extracted from the GRACE data. Note that the Canadian PGR signal (Figure 4.1b) is clearly evident in the total signal (Figure 4.1a).

The fact that this signal can be so clearly identified by simply looking at Figure 4.1a, suggests that GRACE will be able to deliver this signal reasonably free of contamination from the other processes. To obtain a rough estimate of the level of contamination, we truncate the total GRACE result and the PGR signal to the region inside the box. We remove the spatial mean over this region from both signals, and least-squares fit the residual PGR signal to the total signal over this region.

For the parameters used to construct the simulation shown in Figure 4.1a, we obtain a fit parameter of 0.96. When we repeat this procedure for other simulated GRACE data sets, obtained by varying the present-day Antarctic and Greenland ice thickness rates, we obtain fit parameters that vary between about 0.94 and 0.96, which we crudely interpret as implying that GRACE would be able to recover the Canadian PGR signal to an accuracy of about 4–6%. We find we can obtain somewhat better results when we simultaneously fit other parameters to the data (e.g. present-day Antarctic and Greenland mass changes, global sea level change).

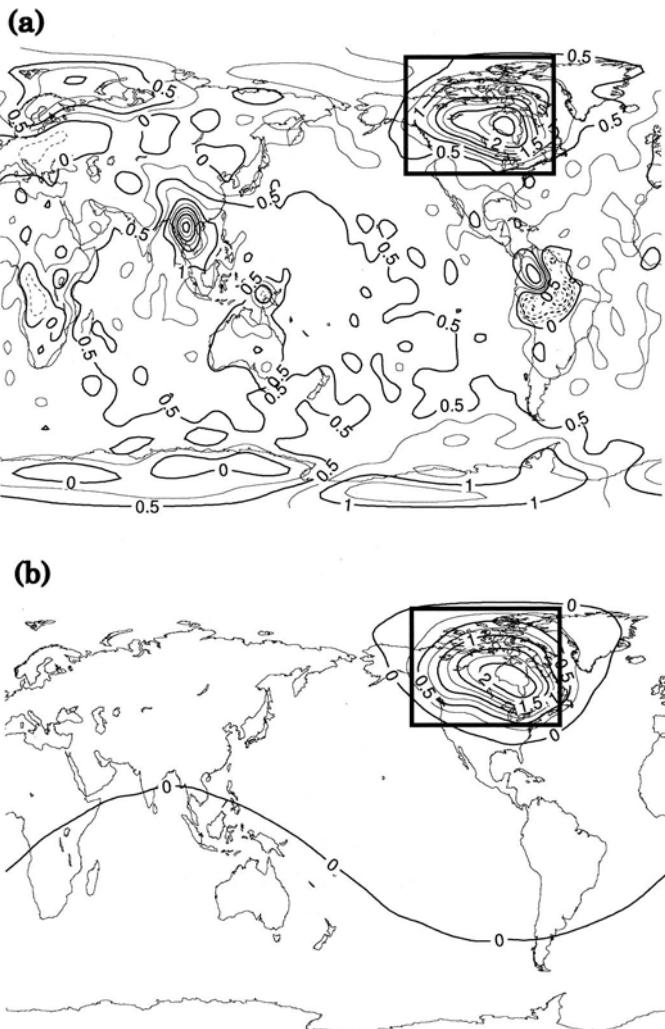


Figure 2. The secular change in the simulated GRACE geoid after five years. Obtained by adding the default PGR signal to predictions of the secular geoid contributions from continental water storage, changes in Greenland and Antarctic ice, the distribution of water in the oceans (including an assumed, non-steric global sea level rise), errors in the GRACE atmospheric pressure corrections, and a realization of the GRACE observational errors. (b) The predicted contribution to the secular geoid change from the default PGR model caused by Canadian ice, alone.

4.2. GRACE RECOVERY OF PGR: METHOD 2

The simple method described in Section 4.1 does not address the larger issue of estimating the improvement we can expect in our knowledge of the Earth's viscosity profile as a result of GRACE. A full resolution of that issue would require a simultaneous inversion of simulated GRACE data and other PGR-related observations,

and should also take into account the likely uncertainties in the ice deglaciation model. Instead, we here adopt the following, more modest approach.

We generate secular gravity fields for a large number of PGR models. For each model we use ice model ICE-3G and the elastic structure given in Table 1, but we vary the lithospheric thickness and the viscosity values in the uppermost mantle, the transition zone, and the lower mantle. We remove these PGR secular gravity fields, one at a time, from the simulated secular GRACE gravity field shown in Figure 4.1a. That simulated field was constructed using our default PGR model. After removing a PGR model, we compute the generalized prediction error, R^2 , defined as:

$$R^2 = \sum_{l=2}^{l_{\max}} \sum_{m=0}^l \frac{[\dot{C}_{lm}^{\text{sim}} - \dot{C}_{lm}^{\text{mod}}]^2 + [\dot{S}_{lm}^{\text{sim}} - \dot{S}_{lm}^{\text{mod}}]^2}{\sigma^2(l)} \left/ \sum_{l=2}^{l_{\max}} \frac{(2l+1)}{\sigma^2(l)} \right. \quad (3)$$

where the $\sigma(l)$ are the degree amplitudes of the secular GRACE measurement errors from Thomas and Watkins, the $\dot{C}_{lm}^{\text{sim}}$ and $\dot{S}_{lm}^{\text{sim}}$ are the Stokes' coefficient rates for the simulated GRACE data, and the $\dot{C}_{lm}^{\text{mod}}$ and $\dot{S}_{lm}^{\text{mod}}$ are the PGR coefficient rates predicted by the Earth model removed from the simulated data. The PGR model that minimizes the generalized prediction error (3) is interpreted as the preferred model. We compare the viscosity values used to construct that preferred model, with the viscosity values of the default model used to generate the simulated data.

After finding R_{\min}^2 , i.e., the generalized prediction error for the model that best fits the measurements, we estimate confidence intervals via the likelihood ratio method (Beck and Arnold, 1977). If the errors are jointly normal, zero mean, and uncorrelated, then the confidence region with the probability α of containing the solution, corresponds to the volume of the model parameter space for which

$$R^2 \leq R_{\min}^2 \left[1 + \frac{M}{n-M} \mathcal{F}_{\alpha}^{-1}(M, n-M) \right], \quad (4)$$

where M is the number of model parameters, n is the number of measurements, and \mathcal{F}^{-1} is the inverse of the F cumulative distribution function.

A complicating factor is that the PGR models include perturbations to the gravity field over Antarctica and Greenland, since ICE-3G includes Holocene deglaciation of those regions. The effects of present-day changes in Antarctic and Greenland ice included in the simulated GRACE data, tend to look similar in spatial pattern to those PGR contributions. Thus, as in Section 4.1, we here address the more restricted problem of focusing on the PGR signal only over Canada. We fit and remove, from $\dot{C}_{lm}^{\text{sim}} - \dot{C}_{lm}^{\text{mod}}$ and $\dot{S}_{lm}^{\text{sim}} - \dot{S}_{lm}^{\text{mod}}$, the secular gravitational effects of uniform ice changes in Antarctica and Greenland, and a uniform non-steric sea level change. We do this prior to constructing R^2 .

The R^2 results are shown in Figure 4.2a-c for PGR models in which we vary pairs of parameters (e.g., uppermost mantle viscosity and transition zone viscosity, etc.) in a grid-search fashion, keeping the remaining parameters fixed to the default

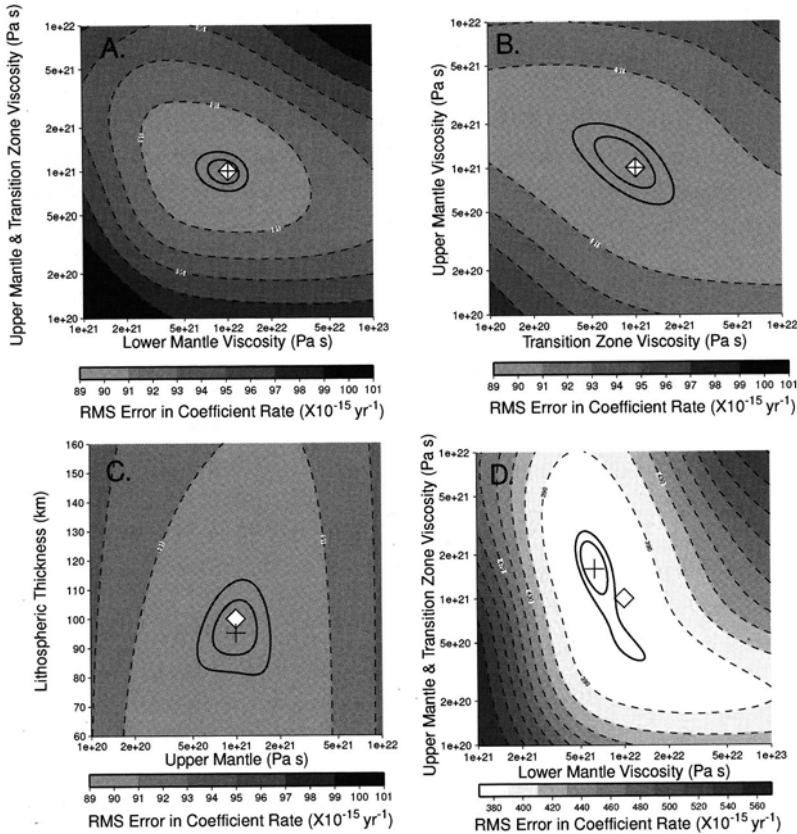


Figure 3. Square root of the generalized prediction error as defined in equation (3), for (a) lower mantle viscosity and uppermost mantle/transition zone viscosity, (b) uppermost mantle and transition zone viscosities, (c) lithospheric thickness and uppermost mantle viscosity. Panel (d) is similar to panel (a), except that a different ice history (ICE-4G) is used to calculate models in the parameter grid search than is used to generate the simulated data (ICE-3G).

values. We include one scenario (Figure 4.2a) in which the uppermost mantle and transition zone viscosities are set to the same value, which is then varied simultaneously with the lower mantle viscosity. Confidence intervals are estimated using $M = 2$ in equation (4). The viscosity values that produce the minimum prediction error are at the centers of the crosses, and the values used for the default model are at the centers of the diamonds. The solid contours represent the 65% and 95% confidence contours.

In general, the values that produce the minimum prediction errors agree well with the default values. The 95% confidence limits on the lower mantle viscosity are between about 6×10^{21} and 14×10^{21} Pa s. Since the default value is 10×10^{21} Pa s, we infer that GRACE is capable of inferring the lower mantle viscosity to about $\pm 40\%$. For the combined uppermost mantle/transition zone shown

in Figure 4.2a, the 95% confidence interval extends between 0.75×10^{21} and 1.3×10^{21} Pa s, which agrees with the default value (1.0×10^{21} Pa s) to within $\pm 30\%$. The results shown in Figure 4.2c suggest that the lithospheric thickness can be recovered to within $\pm 15\text{--}20\%$ (the 95% confidence interval extends between 80 and 115 km). Figure 4.2b shows that GRACE would have a harder time discriminating between separate viscosities in the uppermost mantle and transition zone, where for each of those parameters the 95% confidence interval includes numbers that are between one half and twice the correct value. This is presumably because there is a rough proportionality between vertical and horizontal scales in the PGR solution, so that resolving these two thin layers requires information at short scales. GRACE will not do as well at short scales as at longer scales.

The results shown in Figure 4.2a-c assume there are no errors in the ice model used to calculate the PGR signal. Although this assumption is certainly false, its effects are difficult to determine since there is no obvious way to estimate the ice model errors. To get some idea of how large these effects might be, we consider a second ice history model: ICE-4G from Peltier (1994). ICE-4G is an improved model that uses new constraints which were not available when ICE-3G was developed. We generate the synthetic GRACE data using a PGR signal calculated from ICE-3G, but we use ICE-4G to compute the PGR signals used in the grid search.

The results shown in Figure 4.2d, show that the difference between the ice models can have significant effects on the recovered viscosity. The preferred uppermost mantle/transition zone viscosity is now about 1.6×10^{21} Pa s, or 1.6 times the default value. The preferred lower mantle mantle viscosity is 6×10^{21} Pa s, 40% less than the default value. It is of course possible that this analysis underestimates the effects of true uncertainty in the ice model, since all the data used to constrain ICE-3G were also incorporated into ICE-4G. However, this is perhaps the best estimate we can make for the likely uncertainties.

5. Discussion and Summary

The results presented in Section 3 show that after five years of operation, GRACE should deliver a large number of secular gravity coefficients (in excess of 2000) with errors that are smaller than the PGR signal. In fact, there should be several hundred coefficients with errors smaller than the differences between plausible PGR models. We conclude that the accuracy of the GRACE measurements will be good enough that it is not likely to be a limiting factor for PGR studies.

Instead, the overriding concern is whether the PGR signal can be separated from other sources of secular gravity. This is partly a question of whether GRACE has sufficient resolution to differentiate between the different spatial patterns of the various signals. But even perfect resolution would not allow the separation of every combination of possible signals. For example, separating the PGR signal over Greenland from the combined effects of present-day changes in Greenland

ice mass and secular changes in the surrounding ocean mass distribution, is not possible without invoking additional assumptions about the nature of those signals [see, for example, Wu et al., 2002].

The results presented in Section 4 address the signal separation issue. They suggest that by using GRACE data alone, and if the ice model were perfectly known, it could conceivably be possible to determine the lower mantle viscosity, with 95% confidence, to $\pm 40\%$. Surprisingly, the results also show that GRACE could determine the combined viscosity of the uppermost mantle/transition zone to even better accuracy: $\pm 30\%$. It would be more difficult to use GRACE to solve for separate viscosities in the uppermost mantle and transition zone.

These results do not fully consider the possible effects of unmodeled complexities in the PGR model. When we applied the method described in Section 4.2, we assumed that the set of PGR models removed from the simulated GRACE data included a model with the correct viscosity profile. In reality, the Earth's viscosity profile could well be far more complex, in both its radial and horizontal stratification, than that assumed for any of the forward models constructed for an analysis of this type. The issues then are whether the best-fitting forward model would have a viscosity profile close to the real viscosity profile; and whether it would be possible to converge to the more realistic model by using the sort of minimization method described in Section 4.2.

Errors in the ice model may be even more of a concern. The results shown in Figure 4.2d give some indication of the effects those errors might have on the inferred viscosity. Still, we find it encouraging that the viscosity values we obtain using the wrong ice model, are still within a factor of two of the correct values.

Ultimately, the use of GRACE data to help determine the viscosity and the ice history, would best be done using a formal inversion procedure. The GRACE data would not be used alone, but would be combined with other geological and geodetic measurements of the PGR process. The results of this paper suggest that the GRACE data should add considerable resolving power to any such inversion.

Acknowledgements

We thank Dick Peltier for providing the ICE-4 ice model; Chris Milly and Krista Dunne for providing the output of their water storage model; Brooks Thomas and Mike Watkins for estimates of the GRACE measurement errors; Mery Molenaar and Frank Bryan for estimates of the geoid coefficients due to the ocean; and Bert Vermeersen for useful comments on the manuscript. This work was partially supported by NASA grant NAG5-7703.

References

- Beck, J.V., and Arnold, K.J.: 1977, *Parameter Estimation in Engineering and Science*, 501 pp., Wiley, New York.
- Chao, B.F., and Gross, R.S.: 1987, 'Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes', *Geophys. J. R. Astron. Soc.* **91**, 569–596.
- Cheng, M. K., Shum, C. K., and Tapley, B.: 1997, 'Determination of long-term changes in the Earth's gravity field from satellite laser ranging observations', *J. Geophys. Res.* **102**, 22,377–22,390.
- Church, J.A., Gregory, J.M., et al.: 2001, Changes in sea level, in *Climate Change 2001: The Science of Climate Change: Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, edited by J.T. Houghton, L.G. Meira Filho, B.A. Callander, N. Harris, A. Kattenberg, and K. Maskell, Cambridge Univ. Press, New York, 503–555, in press.
- Dickey, J.O., et al.: 1997, *Satellite Gravity and the Geosphere: Contributions to the Study of the Solid Earth and Its Fluid Envelope*, 112 pp., Natl. Acad. Press, Washington, D.C.
- Dukowicz, J.K., and Smith, R.D.: 1994, 'Implicit free-surface method for the Bryan-Cox-Semtner ocean model', *J. Geophys. Res.* **99**, 7991–8014.
- Dziewonski, A., and Anderson, D.L.: 1981, 'Preliminary reference Earth model', *Phys. Earth Planet. Inter.* **25**, 297–356.
- European Centre for Medium-Range Weather Forecasts: 1995, *The description of the ECMWF/WCRP level III-A global atmospheric data archive*, Data Services publication, Reading, England.
- Kalnay, E., et al.: 1996, 'The NMC/NCAR 40-year reanalysis project', *Bull. Am. Meteorol. Soc.* **77**, 437–471.
- Peltier, W. R.: 1994, 'Ice-age paleotopography', *Science* **265**, 195–201.
- Tushingham, A.M., and Peltier, W.R.: 1991, 'ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of postglacial relative sea level change', *J. Geophys. Res.* **96**, 4497–4523.
- Velicogna, I. and Wahr, J.: 2002a, 'Post Glacial Rebound and Earth's Viscosity Structure From GRACE', *J. Geophys. Res.*, submitted.
- Velicogna, I. and Wahr, J.: 2002b, 'Post-glacial rebound and ice mass balance in Antarctica from ICESat/GLAS, GRACE and GPS measurements', *J. Geophys. Res.*, in press.
- Wahr, J., Molenaar, M., and Bryan, F.: 1998, 'Time-variability of the Earth's gravity field: hydrological and oceanic effects and their possible detection using GRACE', *J. Geophys. Res.* **103**, 30,205–30,230.
- Wahr, J., Wingham, D., and Bentley, C.R.: 2000, 'A method of combining GLAS and GRACE satellite data to constrain Antarctic mass balance', *J. Geophys. Res.* **105**, 16,279–16,294.
- Wu, X., Watkins, M., Kwok, R., Ivins, E., and Wahr, J.: 2002, 'Measuring present-day secular change in Greenland ice mass with future GRACE gravity data', *J. Geophys. Res.*, in press.
- Address for Offprints:* Isabella Velicogna, Department of Physics and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Campus Box 390, Boulder, CO 80309-0390 (e-mail: isabella@grove.Colorado.edu)
- John Wahr, Department of Physics and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Campus Box 390, Boulder, CO 80309-0390 (e-mail: wahr@lemond.Colorado.edu)

MEASURING THE DISTRIBUTION OF OCEAN MASS USING GRACE

R.S. NEREM

*Colorado Center for Astrodynamics Research and Aerospace Engineering Sciences Department,
University of Colorado, Boulder, Colorado, USA (nerem@colorado.edu)*

J.M. WAHR

*Cooperative Institute for Research in Environmental Sciences and the Department of Physics,
University of Colorado, Boulder, Colorado, USA (wahr@lemond.colorado.edu)*

E.W. LEULIETTE

*Colorado Center for Astrodynamics, University of Colorado, Boulder, Colorado, 80309-0431, USA
(eric.leuliette@colorado.edu)*

Received: 4 June 2002; Accepted in final form: 26 September 2002

Abstract. The Gravity Recovery and Climate Experiment (GRACE), which was successfully launched March 17, 2002, has the potential to create a new paradigm in satellite oceanography with an impact perhaps as large as was observed with the arrival of precision satellite altimetry via TOPEX/Poseidon (T/P) in 1992. The simulations presented here suggest that GRACE will be able to monitor non-secular changes in ocean mass on a global basis with a spatial resolution of ~ 500 km and an accuracy of ~ 3 mm water equivalent. It should be possible to recover global mean ocean mass variations to an accuracy of ~ 1 mm, possibly much better if the atmospheric pressure modeling errors can be reduced. We have not considered the possibly significant errors that may arise due to temporal aliasing and secular gravity variations. Secular signals from glacial isostatic adjustment and the melting of polar ice mass are expected to be quite large, and will complicate the recovery of secular ocean mass variations. Nevertheless, GRACE will provide unprecedented insight into the mass components of sea level change, especially when combined with coincident satellite altimeter measurements. Progress on these issues would provide new insight into the response of sea level to climate change.

1. Introduction

Measuring changes in sea level is important for corroborating the predictions of global climate models and for determining the socioeconomic impacts of climate change. TOPEX/Poseidon (T/P) has provided precise estimates of global mean sea level, and global patterns of sea level change in general, since its launch in 1992, and the recently launched Jason-1 satellite will continue those measurements. As shown in Figure 1, T/P has observed an increase in global mean sea level of 2.7 mm/year since its launch, as well as a 15 mm short-term increase in global mean sea level during the 1997–1998 ENSO event (Nerem *et al.*, 1999). Figure 1 shows that the change in global mean sea level during the ENSO event was accompanied by an increase in global mean sea surface temperature.



Space Science Reviews **108:** 331–344, 2003.

© 2003 Kluwer Academic Publishers.

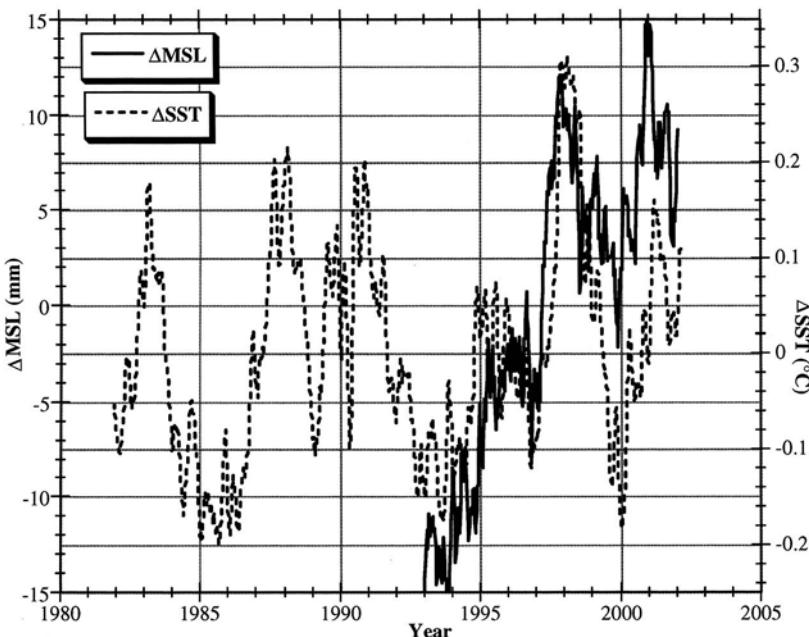


Figure 1. Global mean variations in sea level (from TOPEX/Poseidon) (Nerem *et al.*, 1999) and sea surface temperature (Reynolds and Smith, 1994). Seasonal variations have been removed, and 60-day smoothing has been applied.

While the sea surface temperature results, as well as *in situ* measurements, suggest that much of the sea level change was steric in origin (Cabanes *et al.*, 2001), there is evidence that the global oceans actually lost water mass during the ENSO event, and the steric component was actually double that shown in Figure 1 (Chambers *et al.*, 2000). While the uncertainties in these results are large, they do suggest a complex relationship between mass and steric sea level changes that needs to be better understood if the ultimate cause of the observed sea level change is to be completely unraveled.

The launch of the Gravity Recovery and Climate Experiment (GRACE) mission in 2002 (Tapley and Reigber, 2000) has ushered in a new era in satellite geodesy that has been advocated by numerous scientific groups for many years (Wolff, 1969; Kaula, 1970; Rummel, 1979; Dickey *et al.*, 1997). GRACE consists of two identical satellites separated by ~ 200 km at 450 km altitude and an inclination of 89° . Each satellite carries a precise 3-axis accelerometer, a dual-frequency GPS receiver, and a dual-frequency microwave ranging system that provides precise range and range-rate between the satellites (Thomas, 1999). A number of simulations have demonstrated the ability of GRACE to detect water mass variations in the atmosphere, oceans, and continents (Rodell and Famiglietti, 1999, 2001; Velicogna *et al.*, 2001; Wahr *et al.*, 1998) which will allow unprecedented advances in the fields of atmospheric science, oceanography, and hydrology. Studies have

also shown that GRACE will be able to constrain estimates of polar ice mass balance, thus providing information on the contribution of polar ice to long term sea level change (Velicogna and Wahr, 2002b; Wahr *et al.*, 2000; Wu *et al.*, 2002). In addition, 5 years of GRACE data will allow the detection of mass redistribution on the Earth's surface associated with climate change including the mass component of long-term sea level change (Leuliette *et al.*, 2002), thus providing an important new constraint for global climate models. Finally, 5 years of GRACE data should allow some dramatic improvements in models of glacial isostatic adjustment and mantle viscosity (Velicogna and Wahr, 2002a).

GRACE will allow the estimation of monthly maps of the gravity field which can be used directly to construct maps of changes in ocean bottom pressure with an accuracy of 0.1 mbar over spatial scales of a few hundred km (Wahr *et al.*, 1998; Wunsch *et al.*, 2001). Bottom pressure includes the combined effects of oceanic and atmospheric mass. So, by removing an estimate of atmospheric mass, say from a model that has assimilated meteorological data, the same monthly gravity map can be used to construct a map of changes in the ocean mass distribution. Similarly, the mass component of global sea level change can be determined, if the global mean atmospheric mass change over the ocean is known and removed (Trenberth *et al.*, 1988; Trenberth and Guillemot, 1994). Both the aforementioned studies with T/P data (Chambers *et al.*, 2000; Cabanes *et al.*, 2001) relied on *in situ* oceanographic measurements to compute the steric component of sea level change which could then be subtracted from T/P sea level to determine ocean mass. These *in situ* measurements have very non-uniform coverage in space and time, especially in the southern oceans, and thus large errors can result when trying to make global-scale interpretations. By subtracting the GRACE ocean mass estimate from a map of sea level from T/P, a map of steric sea level change can be computed without using *in situ* measurements, thus GRACE will become a valuable tool for differentiating steric and mass changes in the ocean when combined with satellite altimeter data and *in situ* oceanographic data.

The study of Wahr *et al.* (1998) demonstrated that GRACE should be able to recover estimates of changes in ocean bottom pressure at spatial scales of a few hundred km with roughly monthly temporal resolution and an accuracy of 0.1 mbar or better. While ocean bottom pressure is probably the variable of most interest for assimilation in numerical ocean models, ocean water mass is also a variable of significant interest for many other studies, such as sea level change. This paper presents a simulation of the recovery of ocean mass from GRACE measurements, including an estimate of the expected accuracy of the estimates.

2. Simulation

The following discussion presents the results of a simulation to determine how accurately GRACE can recover variations in ocean water mass. It represents an

extension of the results by Wahr *et al.* (1998), which only described the ocean results from GRACE in terms of ocean bottom pressure (*e.g.* atmosphere plus ocean mass) instead of exclusively ocean mass. The simulation is accomplished by first constructing a “truth” gravity model consisting of monthly modeled mass variations from the oceans and hydrology (the atmosphere is not included because we will subsequently remove it using a model, and add the atmospheric model errors to the measurement errors). Measurement errors (which in this case includes GRACE noise-only errors as well as errors in modeling the atmospheric pressure over the oceans) are added to the “truth” to get the “observed” gravity model. The “truth” and “observed” gravity models are then used to construct surface mass variations (Wahr *et al.*, 1998), which are then differenced to determine the “error” in GRACE-determined ocean mass observations.

2.1. TEMPORAL GRAVITY VARIATIONS

The approach used in this study follows that used in Wahr *et al.* (1998), and thus will not be reproduced in detail here. In summary, the ocean mass redistribution was represented using output from a version of the Parallel Ocean Program (POP) developed at Los Alamos National Laboratory (Dukowicz and Smith, 1994). This “free surface” model is driven by 6 hour surface winds computed by the European Centre for Medium-Range Weather Forecasts (ECMWF) between 1986 and 1996. The model has an average horizontal resolution of 150 km and a vertical resolution varying from 25 m near the surface to 300 m in the deep ocean. The sea surface height, potential temperature, and salinity were sampled every 3 days between January 1, 1987 to January 1, 1996. These results were then combined with an equation of state to estimate density variations in the ocean versus depth, which were in turn integrated over depth to give change in the ocean mass per unit area as a function of latitude and longitude. Previous variants of the POP model have been shown to agree well with TOPEX/POSEIDON altimeter measurements for length scales greater than 300 km and time scales of a few months and longer (Fu and Smith, 1996).

Variations in continental water mass were also modeled, since these variations can leak into the GRACE measurements of ocean mass due to its relatively coarse spatial resolution. Again following Wahr *et al.* (1998), a global gridded data set of soil moisture and snow mass produced by the National Oceanic and Atmospheric Administration Climate Prediction Center (Huang *et al.*, 1996) was used at monthly intervals over the period 1979–1993. The spatial resolution of this model is approximately 200 km. This model does not include water storage variations beneath the soil layer (beneath 1–2 m depth) or any water storage variations in Antarctica.

2.2. GRACE NOISE-ONLY MEASUREMENT ERRORS

The GRACE noise-only measurement errors are modeled as described in Wahr *et al.* (1998). In summary, simulated degree error variances were provided by the

GRACE Project for a 12-day averaging period. The errors are assumed not to be dependent on the spherical harmonic order, m , and they are further assumed to be uncorrelated between different spherical harmonic degrees, l . In addition, the errors are assumed to be independent from one 12-day averaging period to the next, allowing us to compute errors in monthly averages by scaling the 12-day error variances by a factor of 12/30.

The GRACE errors considered here do not include the effects of the mismodeling of ocean tides, also referred to as tidal aliasing (Knudsen and Andersen, 2002; Ray *et al.*, 2001). While these errors are expected to be significant, it is hoped that improved tide models can be developed from the GRACE measurements to minimize this error source, although certain aspects of the tides are unpredictable and may be impossible to model. For the global mean ocean mass estimates, particular attention will have to be given to the S_2 and K_2 solar tides because of their very long alias periods (1 year to secular) due to the slow precession of the GRACE orbit plane.

We have also neglected the more general temporal aliasing errors arising from high frequency variability in the oceans, atmosphere, and continents, which could potentially be a more difficult error source to compensate for (Thompson *et al.*, 2000). Similar errors occur in satellite altimetry (Stammer *et al.*, 2000; Tierney *et al.*, 2000), but they are more difficult to deal with in the GRACE data because they map into gravity in a complex way that is currently not well understood. If no model of the high frequency ocean mass errors is employed, the resulting error due to temporal aliasing may be as large as the GRACE measurement errors themselves (Bettadpur, 2000). Even when using a model to reduce the high frequency variability, temporal aliasing can still be significant because of errors in the models. The current plan by the GRACE Project is that the monthly gravity field estimates will already have models containing high frequency oceanic and atmospheric mass variability removed during the processing of the raw measurements in order to reduce the temporal aliasing. This will be accomplished using a barotropic ocean model forced with ECMWF winds and pressure (Zlotnicki *et al.*, 2001) (thus this model will account for both atmospheric and ocean mass over the ocean), and the ECMWF atmospheric pressure model only over the land. It is expected that the full temporal spectrum of these fields (6 hours) will be removed, and monthly averages of the total mass from the barotropic ocean model (which contains ocean mass plus atmospheric mass over the ocean) will be added back into the final product. Monthly averages of atmospheric mass over the land will not be added back in. Errors in these models are significant, and thus temporal aliasing is still expected to be a significant problem in the final GRACE products.

Finally, we have not included any secular signals or errors in those signals. For example, we have not considered the gravity effects of glacial isostatic adjustment (GIA), or errors from modeling the same. We have also not considered other possible signals, such as the change in the ocean geoid due to the melting of polar ice mass. These signals are likely to be relatively quite large. For example, a reasonable

assessment of errors in current GIA models suggests they may be as large as the secular ocean mass signal itself. Thus, one will have to be particularly careful about differentiating secular and non-secular ocean mass variations. For recovering the former, a simultaneous adjustment which estimates contributions from true ocean mass, GIA, and polar ice mass, may be required. These omissions are not believed to have a significant effect on the non-secular ocean mass error estimates presented here.

2.3. ATMOSPHERIC MODELING ERRORS

Since GRACE will measure the change in the total gravity field of the Earth, we will have to model any changes in the gravity field over the ocean that are not caused by ocean mass variations. We primarily need to be concerned with modeling the redistribution of atmospheric mass, though glacial isostatic adjustment can also contribute a large secular signal. The redistribution of atmospheric mass can be removed from the GRACE monthly gravity estimates using a model, such as those provided by National Centers for Environmental Prediction (NCEP) or European Center for Medium-range Weather Forecasts (ECMWF). However, although these models are driven by real meteorological measurements, the data are quite sparse, especially in Antarctica. Therefore, we must include realistic errors in the adopted atmospheric model in our error estimate for the recovery of ocean mass by GRACE. A commonly used error estimate (Wahr *et al.*, 1998) which we adopt here is the difference between the two model estimates:

$$\delta P(\phi, \lambda) = \frac{P(\phi, \lambda)_{\text{ECMWF}} - P(\phi, \lambda)_{\text{NCEP}}}{\sqrt{2}} \quad (1)$$

The reduction of the errors by root 2 is appropriate if the errors in the two models are independent, which they almost certainly are not. In addition, common errors in the two models will not show up in the difference δP . However, this does give us a reasonable first order assessment of the size of the errors.

If we represent the GRACE monthly estimates as a set of changes in spherical harmonic coefficients ΔC_{lm} , ΔS_{lm} , complete to degree N (where $N = 80$ for our analysis), then the geoid height change as a function of latitude (ϕ) and longitude (λ) may be expressed as a spherical harmonic expansion as:

$$\Delta N(\phi, \lambda) = a_e \sum_{l=0}^N \sum_{m=0}^l P_{lm}(\sin \phi) [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda] \quad (2)$$

where a_e is the equatorial radius of the Earth, and $P_{lm}(\sin \phi)$ are the Legendre polynomials. The ocean mass in units of equivalent water layer thickness can then be expressed as (Wahr *et al.*, 1998):

$$\begin{aligned} \Delta \sigma(\phi, \lambda) &= \frac{a_e \rho_{ave}}{3\rho_w} \sum_{l=0}^N \sum_{m=0}^l \frac{2l+1}{1+k_l} P_{lm}(\sin \phi) \cdot \\ &\quad [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda] \end{aligned} \quad (3)$$

where ρ_{ave} is the mean density of the Earth (5517 kg/m^3), ρ_w is the mean density of water (1000 kg/m^3), and k_l are the Love numbers accounting for an elastic Earth. In addition, a Gaussian weighting function, $W(\phi, \lambda)$, expressed in spherical harmonics (Wahr *et al.*, 1998), W_l , is incorporated to reduce the errors arising from high degree gravity coefficients that may be poorly observed by GRACE. W can also be expressed as a function of the averaging radius, r , around the latitude/longitude location being considered Jekeli (1981). Thus the final functional form for ocean mass estimates given changes in spherical harmonic coefficients from GRACE is:

$$\Delta\bar{\sigma}(\phi, \lambda) = \frac{2\pi a_e \rho_{ave}}{3\rho_w} \sum_{l=0}^N \sum_{m=0}^l \frac{2l+1}{1+k_l} W_l(r) P_{lm}(\sin \phi) \cdot [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda] \quad (4)$$

where the weighting function W is computed as described in Wahr *et al.* (1998) for a given averaging radius r . Thus $\Delta\bar{\sigma}(\phi, \lambda)$ represents an average ocean mass over a spherical cap of radius r centered at latitude (ϕ) and longitude (λ). The simulated recovery of ocean mass using GRACE measurements was accomplished as follows. Monthly “truth” temporal gravity variations were constructed by adding together the spherical harmonic coefficients from the hydrologic and oceanographic models described earlier. The “recovered” gravity variations were constructed by adding spherical harmonics of the pressure model differences to the “truth” spherical harmonics, and then adding GRACE noise-only measurements errors for each spherical harmonic coefficient by taking a random number times the projected GRACE error variance at that degree. Ocean mass distribution for the “truth” and the “observed” spherical harmonics were then computed using Equation (3), using a method for reducing the contamination caused by continental mass variability described by Wahr *et al.* (1998). Errors in the GRACE ocean mass estimates were then computed as:

$$\bar{\varepsilon}(\phi, \lambda) = \Delta\bar{\sigma}(\phi, \lambda)_{\text{observed}} - \Delta\bar{\sigma}(\phi, \lambda)_{\text{truth}} \quad (5)$$

3. Simulation Results

The results of the simulation are illustrated in Figures 2–7. Figure 2 shows the GRACE recovery of ocean mass at a single open ocean location in the Pacific for two different averaging radii — 250 and 550 km. For a 250 km averaging radius, the errors in the GRACE recovery (the difference in the two curves shown) are the same size as the total ocean mass signal at that location. Increasing the averaging radius to 550 km, which reduces errors from more poorly observed higher degree gravity coefficients, but at the expense of spatial resolution, reduces the errors to 2.8 mm RMS water equivalent.

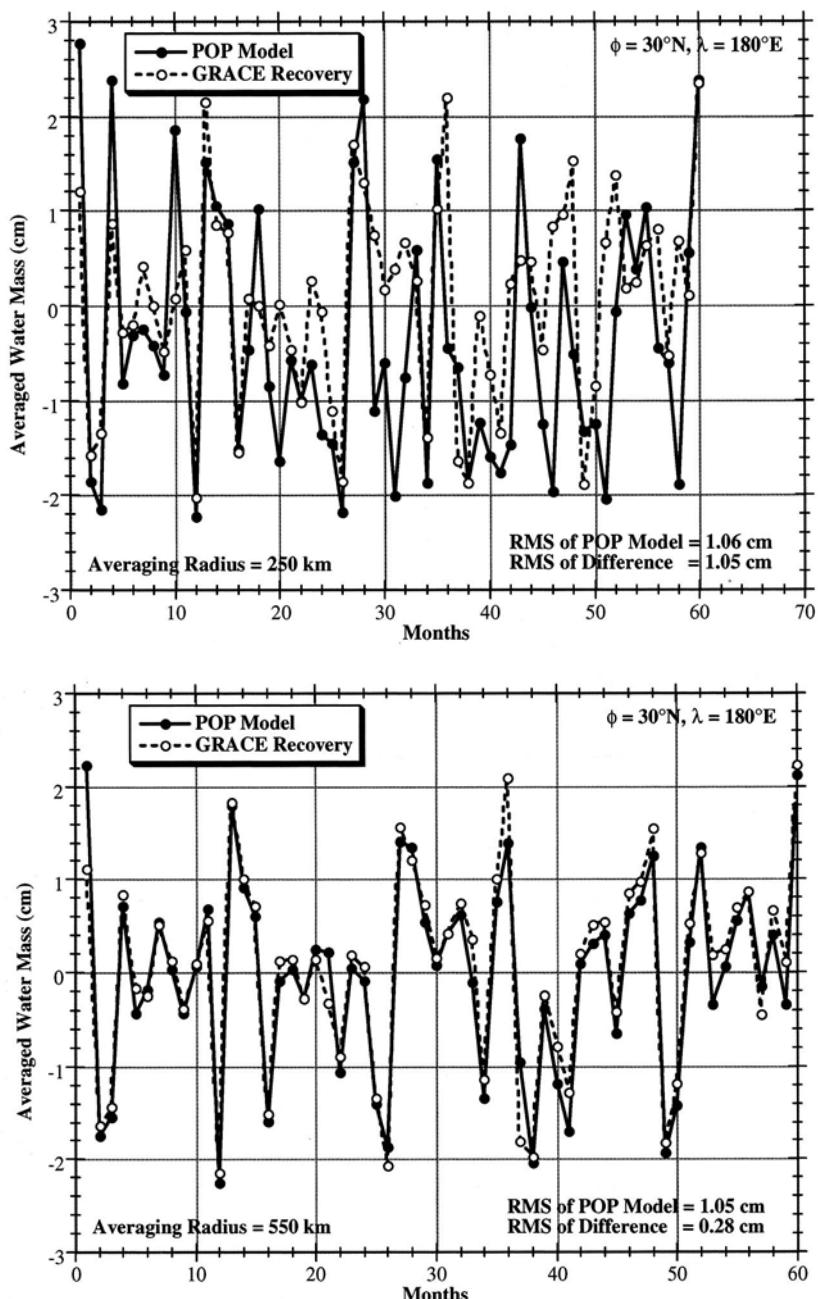


Figure 2. GRACE recovery of ocean mass variations at $\phi = 30^\circ\text{N}$, $\lambda = 180^\circ\text{E}$ in the Pacific for a 250 km averaging radius (top) and a 550 km averaging radius (bottom).

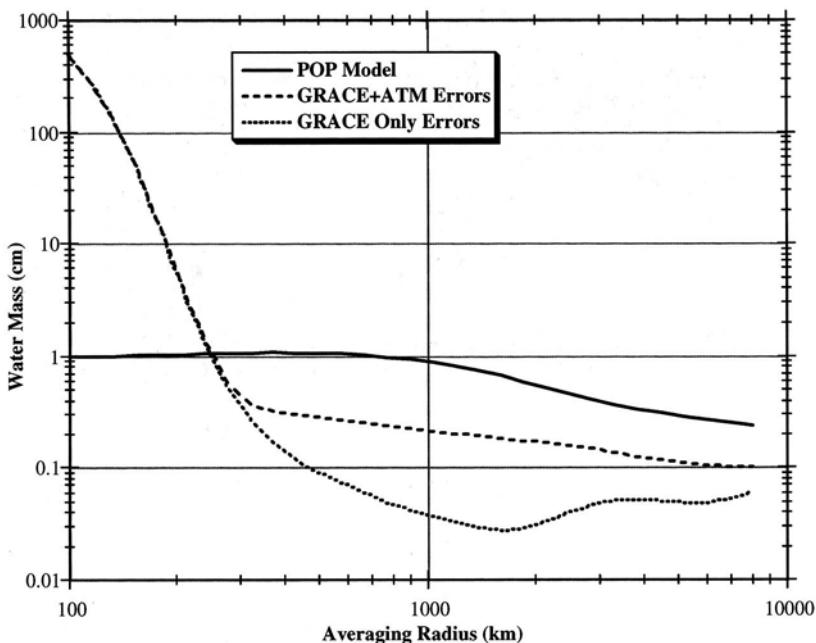


Figure 3. RMS of ocean mass signal and the GRACE ocean mass errors versus averaging radius at $\phi = 30^\circ\text{N}$, $\lambda = 180^\circ\text{E}$ in the Pacific. Two cases for GRACE are shown, one including errors in modeling atmospheric mass and one without (the errors in the latter increase for an averaging radius greater than 1600 km because of leakage of continental mass signal into the ocean mass estimate).

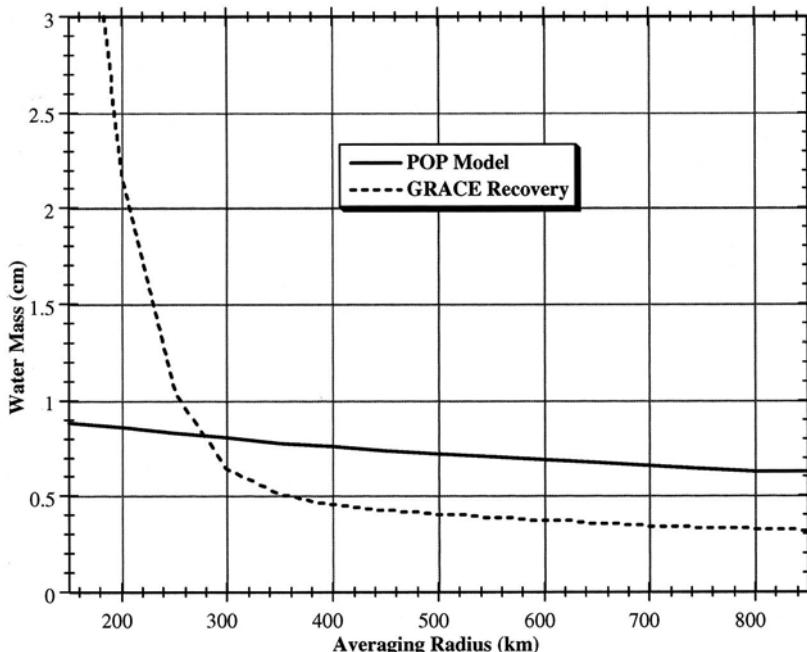


Figure 4. Same as Figure 3, except the errors have been averaged over all ocean locations.

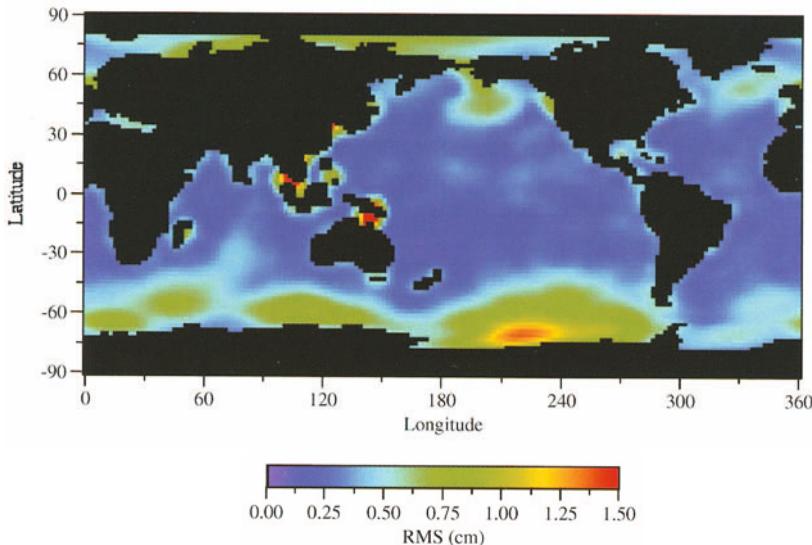


Figure 5. RMS error of the GRACE recovery of ocean mass for a 550 km averaging radius.

Figure 3 shows the same results as Figure 2, but as a function of averaging radius. The signal to noise ratio (the ratio of ocean mass signal to the GRACE ocean mass errors) is about 1 for an averaging radius of 250 km, but a 550 km averaging radius is needed to reduce the SNR to an acceptable level. For this averaging radius, ocean mass may be recovered with an accuracy of about 3 mm. Note that for a 550 km averaging radius, more than half of the error is due to errors in modeling the atmospheric mass, and thus improvements could be gained if improvements in the atmospheric models can be realized. Figure 4 is similar to Figure 3, but now the results have been averaged over all ocean locations.

Figure 5 shows a map of the GRACE recovery of ocean mass for a 550 km averaging radius. Note that the errors are largest around Antarctica due to the larger errors in the atmospheric models there. This is due to a lack of meteorological measurements on the Antarctic continent needed for constraining the models. Figure 6 shows a similar result, but as a function of latitude. Note the larger errors at the higher latitudes, again due to the errors in the atmospheric models.

From Figure 3 it can be seen that for very long averaging radii approaching 10000 km, the error in the ocean mass recovery approaches 1 mm, which suggests that changes in the global mean ocean mass can be recovered to a similar accuracy. In fact, computing a global average of the errors shown in Figure 4 while avoiding the high latitudes (above $\pm 66^\circ$) does result in an error of ~ 1 mm. Only a fraction of this error is due to the GRACE noise-only measurement error (given our assumptions about this error, it reduces dramatically in the global average); most of this error is due to errors in modeling global mean atmospheric mass over the ocean. As an example, Figure 7 shows the difference in atmospheric mass over the oceans

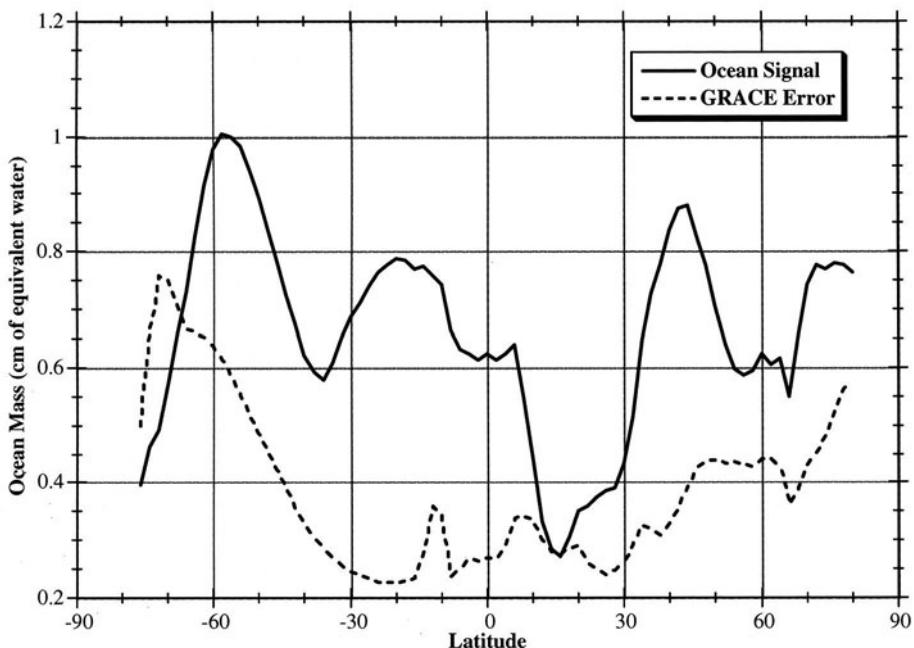


Figure 6. GRACE Recovery of ocean mass versus latitude for a 550 km averaging radius.

(divided by root 2) for the ECMWF and NCEP models. The RMS difference of 1–2 mm (depending on the latitudes used) is approximately the total global ocean water mass error. Thus, with improvements in the atmospheric models, we can expect significant improvements in the recovery of ocean mass using GRACE.

4. Discussion

The simulation presented here demonstrates that GRACE has the potential to recover monthly variations in ocean mass distribution with a spatial resolution of ~ 500 km and an accuracy of ~ 3 mm. In addition, variations in global mean ocean mass may be recovered with an accuracy of ~ 1 mm. A significant fraction of these errors are due to modeling the atmospheric mass redistribution, and thus future improvements in the atmospheric models will significantly improve these results.

The accuracy depicted by these simulations is sufficient to allow the separation of steric and mass components of sea level change when combined with precision satellite altimetry from the TOPEX/Poseidon and Jason-1 missions. It is also sufficient to monitor the contribution of mass variations to global mean sea level change (Figure 1), thus allowing important insight into the causes of sea level change.

A number of error sources have not been considered in this study, notably the aliasing of tide model errors (Knudsen and Andersen, 2002; Ray *et al.*, 2001) and the more general problem of aliasing from high frequency gravity variations. Both

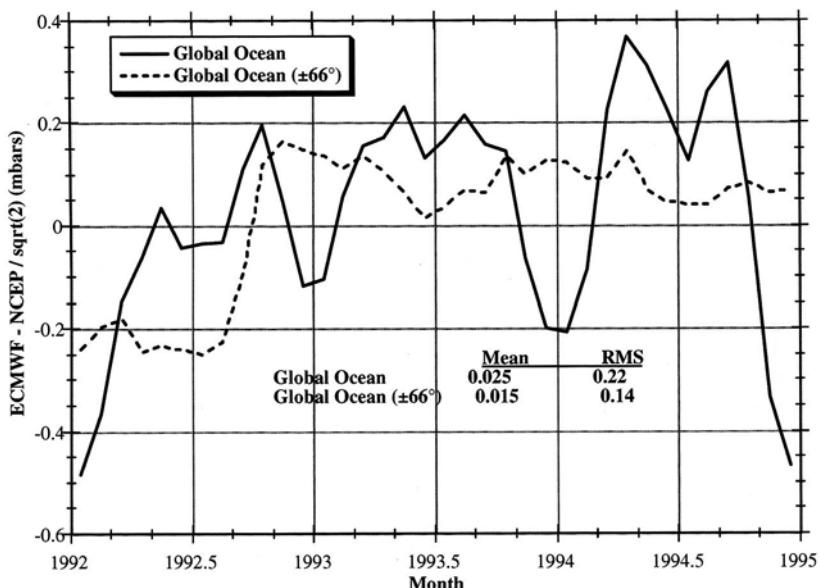


Figure 7. Global mean differences of the ECMWF and NCEP pressure fields (divided by the square root of 2) over the ocean using a 550 km averaging radius.

result from the temporal aliasing of mismodeled gravity variations occurring on time scales much shorter than one month, which is the fundamental temporal resolution of the GRACE data. Tidal aliasing can be reduced through improvements to the tide models using satellite altimeter measurements as well as the actual GRACE measurements. These errors occur at well known astronomical frequencies, and thus is a more tractable problem than aliasing due to more general high frequency variability in the oceans, atmosphere, and on the continents, which is much more difficult to model. Aliasing can also be reduced for future missions by flying multiple satellites or possibly by using a different satellite configuration (Bender *et al.*, 2003).

We have also not considered the contribution of secular signals and error sources, which includes the signals that might be expected from climate change. For example, the recovery of secular variations in global mean ocean mass and their separation from GIA modeling errors was not addressed. This will likely require a simultaneous recovery of ocean mass, GIA errors, polar ice mass contributions, etc., and will require a multi-year time series of GRACE measurements.

References

- Bender, P. L., Nerem, R. S., and Wahr, J. M.: 2003, 'Possible Future Use of Laser Gravity Gradiometers', *Space Sci. Rev.*, this volume.
 Bettadpur, S.: 2000, personal communication.

- Cabanes, C., Cazenave, A., and Le Provost, C.: 2001, 'Sea level change from Topex-Poseidon altimetry for 1993–1999 and possible warming of the southern oceans', *Geophys. Res. Lett.* **28**(1), 9–12.
- Chambers, D. P., Chen, J., Nerem, R. S., and Tapley, B. D.: 2000, 'Interannual mean sea level change and the Earth's water mass budget', *Geophys. Res. Lett.* **27**(19), 3073–3076.
- Dickey, J. O., Bentley, C. R., Bilham, R., Carton, J. A., Eanes, R. J., Herring, T. A., Kaula, W. M., Lagerloef, G. S. E., Rojstaczer, S., Smith, W. H. F., van den Dool, H. M., Wahr, J. M., and Zuber, M. T.: 1997, 'Satellite Gravity and the Geosphere: Contributions to the Study of the Solid Earth and Its Fluid Envelope', National Research Council, Washington, D. C., pp. 112.
- Dukowicz, J. K. and Smith, R. D.: 'Implicit free-surface method for the Bryan-Cox-Semtner ocean model', *J. Geophys. Res.* **99**, 7991–8014.
- Fu, L. and Smith, R. D.: 1996, 'Global ocean circulation from satellite altimetry and high-resolution computer simulation', *Bull. Am. Meteorol. Soc.* **77**, 2625–2636.
- Huang, J., van den Dool, H., and Georgakakos, K. P.: 1996, 'Analysis of model-calculated soil moisture over the United States (1931–1993) and application to long-range temperature forecasts', *J. Clim.* **9**, 1350–1362.
- Jekeli, C.: 1981, 'Alternative methods to smooth the Earth's gravity field', Dept. of Geod. Sci. and Surv., Ohio State Univ., Columbus, Rep. 327.
- Kaula, W. M.: 1970, 'The Terrestrial Environment: Solid Earth and Ocean Physics (The Williamstown Report)', NASA Report CR-1579.
- Knudsen, P. and Andersen, O.: 2002, 'Correcting GRACE gravity fields for ocean tide effects', *Geophys. Res. Lett.* **29**(8), 19-1–19-4.
- Leuliette, E. W., Nerem, R. S., and Russell, G. L.: 2002, 'Detecting Time Variations in Gravity Associated with Climate Change', *J. Geophys. Res.*, 10.1029/2001JB000404.
- Nerem, R. S., Chambers, D. P., Leuliette, E. W., Mitchum, G. T., and Giese, B. S.: 1999, 'Variations in global mean sea level associated with the 1997–1998 ENSO event: Implications for measuring long term sea level change', *Geophys. Res. Lett.* **26**(19), 3005–3008.
- Ray, R. D., Eanes, R. J., Egbert, G. D., and Pavlis, N. K.: 2001, 'Error Spectrum of the Global M2 Ocean Tide', *Geophys. Res. Lett.* **28**, 21–24.
- Reynolds, R. W. and Smith, T. S.: 1994, 'Improved global sea surface temperature analysis', *J. Climate* **7**, 929–948.
- Rodell, M. and Famiglietti, J. S.: 1999, 'Detectability of Variations in Continental Water Storage from Satellite Observations of the Time-Variable Gravity Field', *Wat. Resour. Res.* **35**(9), 2705–2723.
- Rodell, M., and Famiglietti, J. S.: 2001, 'An analysis of terrestrial water storage variations in Illinois with implications for the Gravity Recovery and Climate Experiment (GRACE)', *Water Resour. Res.* **37**(5), 1327–1340.
- Rummel, R.: 1979, 'Determination of the short-wavelength components of the gravity field from satellite-to-satellite tracking or satellite gradiometry: An attempt to an identification of problem areas', *Manuscripta Geodetica* **4**, 107–148.
- Stammer, D., Wunsch, C., and Ponte, R. M.: 2000, 'De-Aliasing of Global High Frequency Barotropic Motions in Altimeter Observations', *Geophys. Res. Lett.* **27**(8), 1175–1178.
- Tapley, B. D. and Reigber, C.: 2000, 'The GRACE Mission: Status and Future Plans', *Eos Trans.* **81**, F311.
- Thomas, J. B.: 1999, 'An Analysis of Gravity-Field Estimation Based on Intersatellite Dual-1-Way Biased Ranging', Jet Propulsion Laboratory, Pasadena, California.
- Thompson, P. F., Bettadpur, S., Kim, J. R., and Watkins, M. M.: 2000, 'Short period variations in the gravity field and their impact on GRACE science', *Eos Trans.* **81**(48), F310.
- Tierney, C., Wahr, J., Bryan, F., and Zlotnicki, V.: 2000, 'Short-period oceanic circulation: implications for satellite altimetry', *Geophys. Res. Lett.* **27**(9), 1255–1258.
- Trenberth, K. E. and Guillemot, C. J.: 1994, 'The total mass of the atmosphere', *J. Geophys. Res.* **99**(D11), 23,079–23,088.

- Trenberth, K. E., Christy, J. R., and Olson, J. G.: 1988, 'Global atmospheric mass, surface pressure, and water vapor variations (abstract)', *J. Geophys. Res.* **93**(D9), 10,925–10,925.
- Velicogna, I. and Wahr, J. M.: 2002a, 'Post glacial rebound and Earth's viscosity structure from GRACE', *J. Geophys. Res.*, in press.
- Velicogna, I. and Wahr, J. M.: 2002b, 'A method for separating Antarctic postglacial rebound and ice mass balance using future ICESat Geoscience Laser Altimeter System, Gravity Recovery and Climate Experiment, and GPS satellite data', *J. Geophys. Res.*, 10.1029/2001JB000708.
- Velicogna, I., Wahr, J. M., and Van den Dool, H.: 2001, 'Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals?', *J. Geophys. Res.* **106**(B8), 16,415–16,434.
- Wahr, J. M., Molenaar, M., and Bryan, F.: 1998, 'Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE', *J. Geophys. Res.* **103**(B12), 30,205–30,230.
- Wahr, J. M., Wingham, D., and Bentley, C.: 2000, 'A method of combining ICESat and GRACE satellite data to constrain Antarctic mass balance', *J. Geophys. Res.* **105**(B7), 16,279–16,294.
- Wolff, M.: 1969, 'Direct Measurements of the Earth's Gravitational Potential Using a Satellite Pair', *J. Geophys. Res.* **74** (22), 5295–5300.
- Wu, X., Watkins, M. M., Ivins, E. R., Kwok, R., Wang, P., Wahr, J. M.: 2002, 'Toward global inverse solutions for current and past ice mass variations: Contribution of secular satellite gravity and topography change measurements', *J. Geophys. Res.*, 10.1029/2001JB000543.
- Wunsch, J., Thomas, M., and Gruber, T.: 2001, 'Simulation of oceanic bottom pressure for gravity space missions', *Geophys. J. Int.* **147**, 428–434.
- Zlotnicki, V., Ali, A. H., and Ponte, R.: 2001, 'Atmospheric forcing of a barotropic ocean model to dealias altimetry and GRACE', *Eos Trans.* **82**(47), F288.

MONITORING CHANGES IN CONTINENTAL WATER STORAGE WITH GRACE

SEAN SWENSON and JOHN WAHR

*Department of Physics and Cooperative Institute for Research in Environmental Sciences,
University of Colorado, Boulder, Colorado, USA*

Received: 3 September 2002; Accepted in final form: 13 December 2002

Abstract. The Gravity Recovery and Climate Experiment, GRACE , will enable the recovery of monthly estimates of changes in water storage, on land and in the ocean, averaged over arbitrary regions having length scales of a few hundred km and larger. These data will allow the examination of changes in the distribution of water in the ocean, in snow and ice on polar ice sheets, and in continental water and snow storage. Extracting changes in water storage from the GRACE dataset requires the use of averaging kernels which can isolate a particular region. To estimate the accuracy to which continental water storage changes in a few representative regions may be recovered, we construct a synthetic GRACE dataset from global, gridded models of surface-mass variability. We find that regional changes in water storage can be recovered with rms error less than 1 cm of equivalent water thickness, for regions having areas of 4×10^5 km² and larger. Signals in smaller regions may also be recovered; however, interpretations of such results require a careful consideration of model resolution, as well as the nature of the averaging kernel.

1. Introduction

Time-variable gravity changes are caused by a combination of postglacial rebound, fluctuations in atmospheric mass, and the redistribution of water, snow, and ice on land and in the ocean. GRACE, launched in March, 2002, will provide time-variable gravity measurements which should greatly improve upon the current best models. In addition, GRACE will deliver a new estimate of the spherical harmonic coefficients describing the Earth's gravity field approximately every 30 days.

From the gravity field estimates, we expect to infer time-variable changes in mass, averaged over arbitrary regions having length scales of a few hundred km and larger, to accuracies of better than 1 cm of equivalent water thickness. These data will be useful for examining changes in the distribution of water in the ocean, in snow and ice on polar ice sheets, and in continental water and snow storage. These quantities can then be used to assess and improve climate models, to better understand large-scale hydrological processes, and to monitor the distribution of land-based water for agricultural and water resource applications. Combined with radar altimetry over the oceans, these data can improve estimates of the time-varying ocean heat storage, as well as deep ocean currents. In polar regions, GRACE data can be used to study postglacial rebound and, in conjunction with radar and



laser altimetry (ERS-2 RA; Envisat RA-2, IceSat and CryoSat), to constrain the mass-balance of ice sheets.

2. The GRACE Dataset

The geoid, N , can be expressed as a sum of normalized associated Legendre functions, \tilde{P}_{lm} ,

$$N(\theta, \phi) = a \sum_{l=0}^{\infty} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) \{C_{lm} \cos m\phi + S_{lm} \sin m\phi\}, \quad (1)$$

where θ is co-latitude, ϕ is longitude, a is the mean radius of the Earth, and C_{lm} and S_{lm} are dimensionless Stokes coefficients. GRACE will deliver Stokes coefficients up to degree $l \sim 100$ every 30 days. The spatial scale, λ , associated with a Stokes coefficient of degree l can be found approximately by the relation $\lambda = 20,000 \text{ km}/l$; $l \leq 100$ therefore correspond to length scales of about 200 km and larger.

Using these coefficients, it will be possible to infer changes in water storage from one 30-day period to the next. GRACE will be accurate enough to be sensitive to changes in the Earth's gravity field caused by fluctuations in continental water storage and the polar ice sheets, as well as by changes in atmospheric and oceanic mass distribution. The contributions from the atmosphere can be estimated from independent atmospheric data and largely removed (Velocogna et al., 2001; Swenson and Wahr, 2002). Most of the remaining time-variable mass changes are confined to a thin layer at the Earth's surface, so that the vertically integrated water and ice mass can be viewed as a surface mass density.

A local change in water storage, $\Delta\sigma(\theta, \phi)$, can be related to changes in the Stokes coefficients, ΔC_{lm} and ΔS_{lm} , by

$$\Delta\sigma(\theta, \phi) = \frac{a \rho_E}{3} \sum_{l=0}^{\infty} \sum_{m=0}^l \frac{(2l+1)}{(1+k_l)} \tilde{P}_{lm}(\cos \theta) \times \{\Delta C_{lm} \cos m\phi + \Delta S_{lm} \sin m\phi\}, \quad (2)$$

where ρ_E is the average density of the solid Earth, and k_l are the load love numbers representing the effects of the Earth's response to surface loads (Wahr et al., 1998).

The Stokes coefficients produced by GRACE will contain measurement errors. In this paper we use preliminary measurement error estimates provided by B. Thomas and M. Watkins at JPL, that are consistent with those described in the *GRACE Science and Mission Requirements Document* (2001). For values of l greater than about 5, the estimated degree amplitudes of the GRACE satellite errors increase rapidly with increasing l .

3. Spatial Averaging to Improve Accuracy

Spatial averaging reduces the contributions from short wavelength components of both the water storage signal and satellite measurement error. This is beneficial because GRACE will not provide information for wavelengths shorter than about 200 km, and the shortest components that are provided will contain the greatest measurement errors.

The change in vertically integrated water storage averaged over an arbitrary region is

$$\overline{\Delta\sigma}_{\text{region}} = \frac{1}{\Omega_{\text{region}}} \int \Delta\sigma(\theta, \phi) W(\theta, \phi) d\Omega , \quad (3)$$

where Ω_{region} is the angular area of the region, and $W(\theta, \phi)$ is an averaging kernel. If the averaging kernel is expressed as

$$W(\theta, \phi) = \frac{1}{4\pi} \sum_{l=0}^{\infty} \sum_{m=0}^l \tilde{P}_{lm}(\cos \theta) \{W_{lm}^c \cos m\phi + W_{lm}^s \sin m\phi\}, \quad (4)$$

the regional average becomes

$$\overline{\Delta\sigma}_{\text{region}} = \frac{a \rho_E}{3 \Omega_{\text{region}}} \sum_{l=0}^{\infty} \sum_{m=0}^l \frac{(2l+1)}{(1+k_l)} (W_{lm}^c \Delta C_{lm} + W_{lm}^s \Delta S_{lm}) . \quad (5)$$

The exact averaging kernel is a function which describes the shape of the region according to

$$W(\theta, \phi) = \begin{cases} 0 & \text{outside the basin,} \\ 1 & \text{inside the basin.} \end{cases} \quad (6)$$

The exact averaging kernel changes discontinuously from a value of 1 to 0 at the basin boundaries, resulting in W_{lm}^c and W_{lm}^s which have relatively large amplitudes at short-wavelengths. These short-wavelength coefficients lead to the possible inclusion of the Stokes coefficients which contain the largest satellite measurement errors. Measurement error can be reduced by creating an averaging kernel which varies smoothly across its boundary, because this approximate averaging kernel has less power at short-wavelengths than does the exact averaging kernel. The disadvantage of using an approximate averaging kernel is that $\overline{\Delta\sigma}_{\text{region}}$ will be influenced by mass signals outside the region, as well as over- or under-estimating the contribution of the signal inside the basin. We refer to this error as leakage. However, the tradeoff between increasing leakage error and decreasing measurement error can be quite favorable. We will show that approximate averaging kernels can be constructed which drastically reduce satellite measurement error while producing a relatively small increase in leakage error.

4. Approximate Averaging Kernels

There are three primary error sources to consider when attempting to recover the water storage signal from GRACE data: satellite measurement error, leakage error, and atmospheric pressure error (the error due to imperfectly removing the atmospheric component of the gravity field). While pressure errors are for the most part independent of kernel size and shape, measurement and leakage errors depend greatly on the exact shape of the averaging kernel. Swenson and Wahr (2002) describe minimization techniques which incorporate measures of satellite and leakage errors to create an optimal averaging kernel. These methods achieve one of three goals: minimizing the sum of the satellite and leakage errors, fixing measurement error to a specific value and minimizing leakage error, or fixing leakage error to a specific value and minimizing measurement error.

5. Synthetic GRACE Data

Wahr et al. (1998) describe the process of creating synthetic GRACE gravity data, which is summarized here. Models of surface-mass variability were used to create five years of monthly GRACE gravity solutions, expressed as a set of Stokes coefficients for each month. The water storage signal came from a version of the Land Dynamics model (Milly and Shmakin, 2001) run at the Geophysical Fluid Dynamics Laboratory. Groundwater, root-zone, and snowpack store components are modeled at 1 degree resolution. The contribution from the ocean comes from a model derived from the Los Alamos Parallel Ocean Program ocean general circulation model (Dukowicz and Smith, 1994), run at the National Center for Atmospheric Research. Errors in the process of removing the atmospheric gravity signal were estimated as the difference between the ECMWF and NCEP global gridded surface pressure fields, divided by $\sqrt{2}$. Satellite measurement errors, as described in section 2, were also included.

6. Assessment of GRACE Recovery

We estimate the accuracy with which continental water storage signals may be recovered from GRACE data by constructing regional hydrology estimates from our simulated GRACE dataset. The simulated recovery can be compared with the correct regional averages obtained from the hydrology model alone. The differences are a measure of the error in the GRACE estimates.

Figure 1 (a) shows the result of using the exact averaging kernel to estimate the water storage signal of the Mississippi river basin. The true hydrology signal, computed directly from the model, is shown as a solid line. The recovered signal, derived from the synthetic GRACE dataset, is represented by a dashed line. The

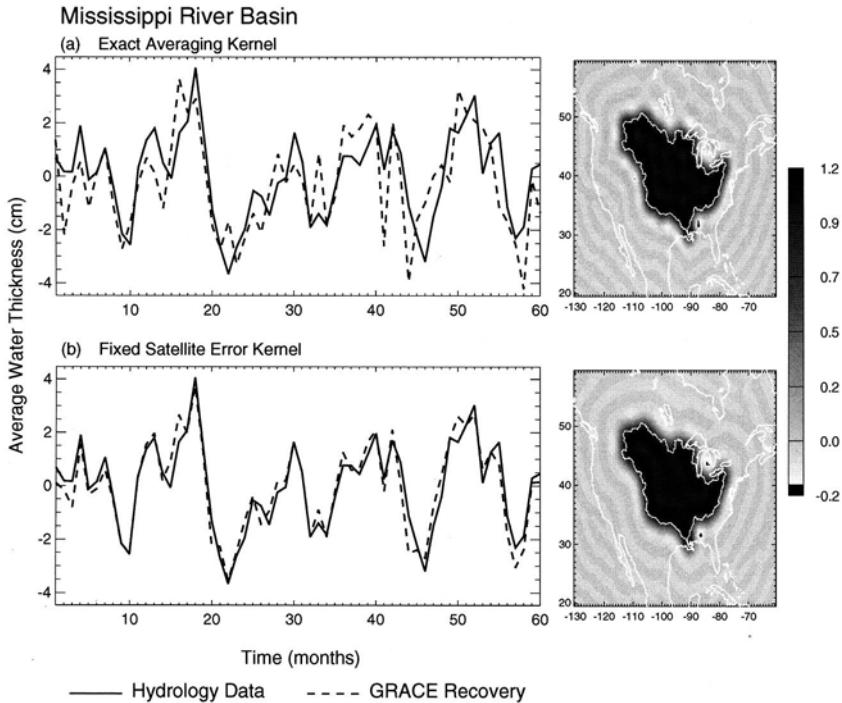


Figure 1. Comparison of synthetic GRACE recoveries for exact and approximate averaging kernels. (a) left panel shows the recovered time series for an exact averaging kernel of the Mississippi river basin. Right panel shows the shape of the exact averaging kernel. (b) left panel shows the time series for an approximate averaging kernel designed to produce 0.5 cm of satellite measurement error. Right panel shows the shape of the approximate averaging kernel.

RMS of the true signal is 1.63 cm, while the RMS of the total difference is 1.27 cm (Table I). Leakage error is 0.02 cm; it is not zero because GRACE only provides Stokes coefficients to degree 100. The main component of error is measurement error. Figure 1 (b) shows the results of using the approximate averaging kernel, which is designed to produce 0.5 cm of satellite error, while minimizing the leakage error. The resulting error estimates are shown in Table I. While leakage error is nearly the same as that obtained using the exact averaging kernel, satellite error is reduced to roughly one third the value in the previous case. The total error is now only 0.5 cm.

As basin size decreases, the amplitude of short-wavelength components in the averaging kernel generally increase, leading to the inclusion of larger measurement errors. Figure 2 shows the results of using the approximate averaging kernel to recover the water storage signal while constraining measurement error to 0.5 cm, for the Ohio and Red river basins. The true hydrology signals are 3.17 and 4.08 cm respectively (Table II). While measurement errors again dominate the total

TABLE I

Mississippi River Basin	Area $3.89 \times 10^5 \text{ km}^2$	
True Signal	1.63 cm	
RMS Basin Average Recovery Errors (cm)		
	Exact	Fixed Measurement Error
Recovered Signal	1.82	1.69
Total Error	1.27	0.50
Satellite Error	1.24	0.42
Hydrology Leakage	0.02	0.02
Atmospheric Error	0.26	0.26
Ocean Leakage	0.02	0.02

error budget, leakage errors are relatively larger for these basins. In order to allow only 0.5 cm of measurement error, the approximate averaging kernel of a small basin must be relatively smoother than that of a large basin. Thus, the approximate averaging kernel of the larger Ohio river basin, shown in the right panel of figure 2 (a), looks more like its exact averaging kernel than does the approximate averaging kernel of the smaller Red river basin, shown in the right panel of figure 2 (b). Because of the smoothness constraints, the Red averaging kernel has a maximum value of 0.8, rather than 1, and the non-zero portion of the averaging kernel extends further from the basin boundary. Still, the monthly water storage variability in each of these basins can be recovered to better than 1 cm RMS.

We use the Iowa river basin ($\text{area} = 8.23 \times 10^4 \text{ km}^2$) as an example of the caveats that must be considered when using synthetic GRACE data to estimate the accuracy of an actual GRACE recovery for a small region. Figure 3 (a) shows the recovery made with an averaging kernel designed to have a leakage ratio, R , of 0.5. The leakage ratio is defined as the spatial variance of the difference between the exact and approximate averaging kernels, normalized by the spatial variance of the exact averaging kernel (Swenson and Wahr, 2002). A small R signifies a kernel which closely approximates the exact kernel, while a large R signifies a kernel that is relatively smoother. In this case, a value of $R = 0.5$ creates a kernel which looks much like the exact averaging kernel. The total error budget, which is dominated by satellite measurement error, overwhelms the signal (Table III). To reduce the error, we created an averaging kernel constrained to include only 0.5 cm of measurement error. Figure 3 (b) shows the recovery made with this kernel. The total error drops from 11.29 cm to 1.06 cm. However, an examination of the averaging kernel shows that it has been drastically smoothed, and samples a much greater region than that of the exact kernel. The reason for the relatively small increase in leakage error

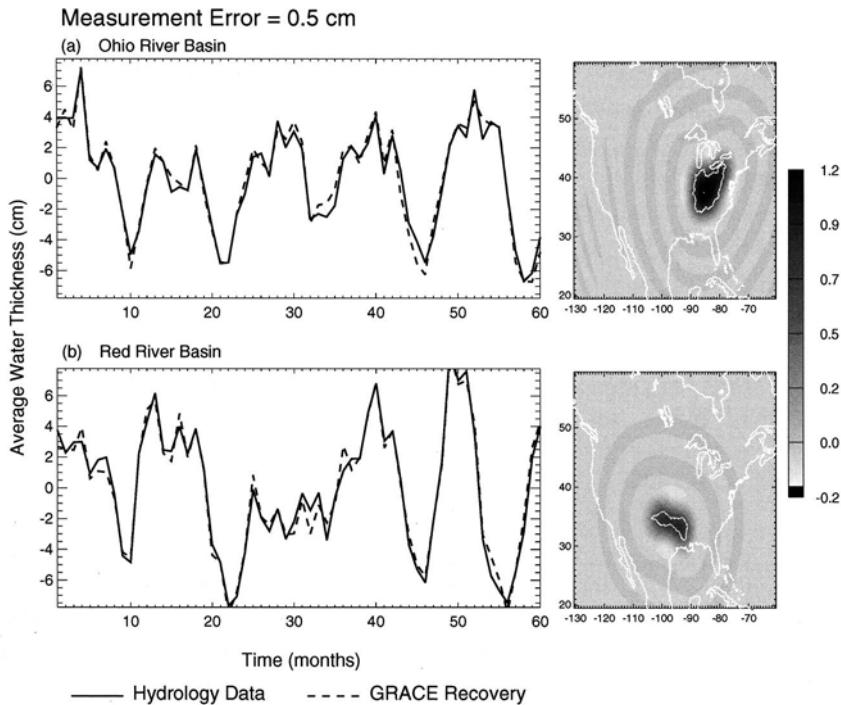


Figure 2. Synthetic GRACE recoveries with approximate averaging kernels for (a) the Ohio river basin, and (b) the Red river basin. Both approximate averaging kernels are designed to produce 0.5 cm of satellite measurement error.

TABLE II

Satellite Measurement = 0.5 cm		
RMS Basin Average Recovery Errors (cm)		
Basin	Ohio	Red
Area ($\times 10^5$ km 2)	7.36	4.01
True Signal	3.17	4.08
Recovered Signal	3.34	4.01
Total Error	0.59	0.61
Satellite Error	0.56	0.54
Hydrology Leakage	0.11	0.27
Atmospheric Error	0.21	0.22
Ocean Leakage	0.01	0.03

TABLE III

Iowa River Basin	Area	$8.23 \times 10^4 \text{ km}^2$	
True Signal		2.80 cm	
RMS Basin Average Recovery Errors (cm)			
	Fixed leakage ratio	Fixed satellite error	Fixed LR with LI
Recovered Signal	12.24	2.35	2.82
Total Error	11.29	1.06	0.26
Satellite Error	11.27	0.53	0.01
Hydrology Leakage	0.23	0.85	0.24
Atmospheric Error	0.19	0.19	0.20
Ocean Leakage	0.03	0.03	0.03

is that the water storage signal in the hydrology model we are using to construct our synthetic GRACE data is correlated over distances that are long compared to the length scale of the basin. Studies of the spatial scales of soil moisture correlations have shown two different regimes of variability: a small-scale correlation on the order of a kilometer or less, and a large-scale correlation on the order of hundreds of kilometers (Entin et al., 2000). Short-wavelength components of the water storage signal are not resolved by most global-scale hydrology models, and therefore we cannot determine if their contribution is important in a region of this size. Thus, the choice to employ a smooth averaging kernel such as that shown in Figure 3(b), rather than a more compact kernel such as that shown in 3 (a), should partly depend on whether the user believes the water storage variability within the basin of interest is well correlated with the variability in the surrounding region.

A possible improvement to the current GRACE mission is the replacement of the microwave ranging system with a laser interferometer. Measurement errors with this system are estimated to be approximately 3 orders of magnitude smaller than those obtained from GRACE (Watkins et al., 1999). Figure 3 (c) shows a recovery made using synthetic GRACE data which incorporate these measurement error estimates. An approximate averaging kernel with $R = 0.5$ now leads to a recovered water storage signal that has an RMS of 2.82 cm and a total error of 0.26 cm, of which measurement error contributes a negligible amount.

7. Summary

Monthly averages of Stokes coefficients provided by GRACE will allow the estimation of regional averages of changes in continental water storage. One way to estimate the accuracy of an actual GRACE recovery is to generate simulated

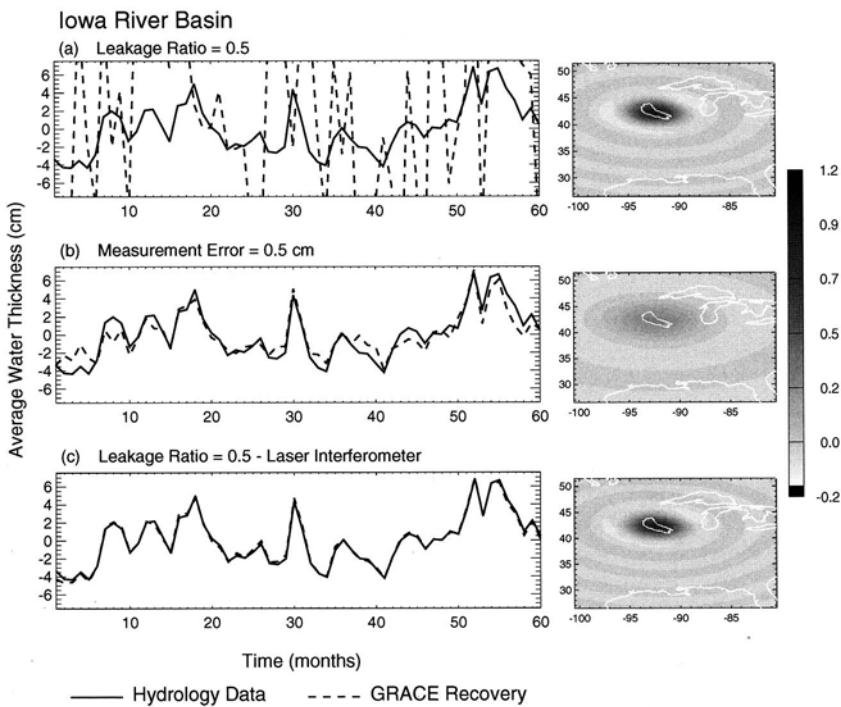


Figure 3. Comparison of synthetic GRACE recoveries for different approximate averaging kernels of the Iowa river basin. (a) recovery using a fixed leakage ratio averaging kernel, $R = 0.5$. (b) fixed satellite measurement error, error = 0.5 cm. (c) fixed leakage ratio with Laser Interferometer measurement error estimates.

recoveries from synthetic GRACE data. Our simulations show that using an exact averaging kernel leads to the inclusion of large satellite measurement errors. To reduce measurement errors to an acceptable level, approximate averaging kernels must be used when estimating regional averages. While reductions in measurement error occur at the expense of sampling signals from surrounding areas, the contribution from this leakage error is typically small for regions having areas in excess of $4 \times 10^5 \text{ km}^2$. Errors due to the imperfect removal of the atmospheric gravity signal from GRACE data account for about 0.20 to 0.26 cm in the basins examined here; these values are consistent with that obtained by Schrama (this issue). Recovery errors of less than 1 cm of water thickness appear to be obtainable for basins of $4 \times 10^5 \text{ km}^2$ and larger. As the size of the region under study decreases, its averaging kernel must become increasingly smooth. As a kernel becomes smoother, it samples more of the signal from areas outside the region of interest. When the length scale of the region approaches the resolution of the models surface-mass variability used to construct the synthetic GRACE dataset, the lack of short-wavelength signal

components likely leads to overly optimistic estimates of the accuracy of GRACE recoveries.

Acknowledgements

This work was partially supported by NASA grants NAG5-9450 and NAG5-7703 to the University of Colorado, by a Cires Innovative Research Grant, and a NASA Earth Systems Science Fellowship awarded to Sean Swenson.

References

- Dukowicz, J.K., and R.D. Smith: 1994, Implicit free-surface method for the Bryan-Cox-Semtner ocean model, *J. Geophys. Res.* **99**, 7991–8014.
- Entin, J., A. Robock, K. Vinnikov, S. Hollinger, S. Liu, and A. Namkhai: 2000, Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.* **105**(D9), 11,865–11,877.
- GRACE Science and Mission Requirements Document*: 2001, GRACE 327-200, JPL Publ. D-15928, Rev. D.
- Milly, P.C.D., and A.B. Shmakin: 2001, Global modeling of land water and energy balances: 1. The Land Dynamics (LaD) Model, submitted to *Journal of Hydrometeorology*.
- Swenson, S., and J. Wahr: 2002, Methods for inferring regional surface-mass anomalies from GRACE measurements of time-variable gravity, *J. Geophys. Res.* **107**(B9), 2193, doi:10.1029/2001JB000576.
- Swenson, S., and J. Wahr: 2002, Estimated effects of the vertical structure of atmospheric mass on the time-variable geoid, *J. Geophys. Res.* **107**(B9), 2194, doi:10.1029/2000JB000024.
- Velicogna, I., J. Wahr, and H. van den Dool: 2001, Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals?, *J. Geophys. Res.* **106**(B8), 16,415–16,434.
- Wahr, J., M. Molenaar, and F. Bryan: 1998, Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.* **103** (B12), 30,205–30,229.
- Watkins, M. M., W.M. Folkner, B. F. Chao, and B. D. Tapley: 1999, The Proposed NASA EX-5 Mission: A Laser Interferometry Successor to GRACE, *Eos Trans. AGU* **80**(46), Nov. 16, 1999.

VI: FUTURE CONCEPTS

ATTITUDE AND DRAG CONTROL: AN APPLICATION TO THE GOCE SATELLITE

E. CANUTO (canuto@polito.it)

*Politecnico di Torino, Dipartimento di Automatica e Informatica, Corso Duca degli Abruzzi 24,
10129 Torino, Italy*

P. MARTELLA (pmartell@to.alespazio.it) and G. SECHI
(gsechi@to.alespazio.it)
Alenia Spazio, Corso Marche 31, Torino, Italy

Received: 23 May 2002; Accepted in final form: 24 July 2002

Abstract. The Gravity field and steady-state Ocean Circulation Explorer (GOCE) satellite, currently planned to be launched in the course of 2006, will require a precise drag compensation and a fine attitude control along the Local Orbiting Reference Frame (LORF) of a polar Sun-synchronous low orbit, allowing the Earth gravity field to be recovered with unprecedented accuracy by post-processing the scientific telemetry. To this aim, the spectral density of the spacecraft linear and angular accelerations must be limited below $0.025 \frac{\mu\text{m}}{\text{s}^2\sqrt{\text{Hz}}}$ and $0.015 \frac{\mu\text{rad}}{\text{s}^2\sqrt{\text{Hz}}}$ respectively, in the frequency range from 5 mHz to 0.1 Hz, the gradiometer measurement bandwidth. In the same range, the orientation errors of the spacecraft in the LORF and of the LORF in the inertial frame must be kept below $10 \frac{\mu\text{rad}}{\sqrt{\text{Hz}}}$. The Drag-Free Mode, encharged of drag-free and attitude control (DFAC) during measurement phases, determines the spacecraft state vector using a very precise gradiometer, one large Field-of-View Star Tracker and a Satellite-to-Satellite Tracking Instrument. Force and torque commands are actuated by two assemblies of thrusters: a single ion-thruster acting along the orbital direction, a set of eight micro-thrusters acting along the other five degrees of freedom. To cover every mission scenario, other control modes have been studied and designed: the Coarse Pointing Mode dedicated to rate damping and Sun acquisition, the Fine Pointing Mode handling the transition to Drag-Free Mode and the Ultimate Safe Mode, a survival operative mode to improve mission reliability. Results presented in this paper give a positive perspective on the solidity of the current DFAC design.

1. Introduction

The Gravity field and steady-state Ocean Circulation Explorer (GOCE) satellite has in charge of improving gravity field and geoid models, essential in a wide range of applications: geodesy, oceanography, solid-Earth physics and others, by collecting measurements with high spatial resolution and uniform accuracy. The objective of the mission is to produce the Earth potential spherical harmonic coefficients with a half-wavelength resolution of 100 km or better, enabling the recovery of geoid heights and gravity anomalies with accuracies better than 1 cm and $10 \frac{\mu\text{m}}{\text{s}^2}$, respectively.



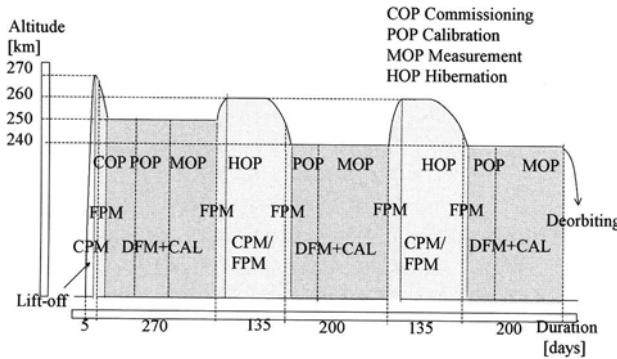


Figure 1. Mission profile showing different phases and relevant drag-free and attitude control (DFAC) modes: ordinates: altitude [km], abscissae: duration [days].

The traditional techniques of gravity field determination have reached their intrinsic limits (Johanssen, 1999). Any advances must rely on space techniques because only they provide global, regular and dense data sets of high and homogeneous quality. GOCE will implement three main concepts for recovering gravity field.

- *Precise orbit determination (POD)* by satellite-to-satellite tracking (SST). SST technique is limited by progressive attenuation of the gravitational field at satellite altitudes, which prohibits the attainment of high spatial resolution.
- *Satellite gravity gradiometry*. An on-board gradiometer (gradio for short) measures the components of the gravity gradient tensor exploiting the classical differential approach for enlightening the effect of small-scale features.
- *Drag-free and attitude control, DFAC*. To extract gravitational field components from orbit and gradiometer measurements, non-gravitational forces must be accurately compensated by a drag-free control mechanism and spacecraft attitude must be accurately aligned to the Local Orbital Reference Frame (LORF), to which gravity measurements will be referred.

Satellite gradiometry and POD by SST tracking are complementary. By means of POD it is possible to reconstruct with high accuracy the lower harmonics of the gravity field, while gradiometry provide better performances at medium and high degrees. The intersection is somewhere close to degree 15. This leads to the definition of a Measurement Bandwidth (MBW) for the gradiometer measurements, i.e. the frequency region from 5 to 100 mHz where the measurement accuracy of the gravity gradient has to be maximized. The GOCE satellite will fly on a circular polar (96.5°) Sun-synchronous orbit at an average altitude between 240 and 270 km. The launch is planned to take place in the course of 2006 and the mission will last about 31 months. The gradiometric measurements are planned to be collected in three phases (MOP) of 6 months each. Main mission phases and the relevant DFAC nominal modes, to be explained below, are shown in Figure 1.

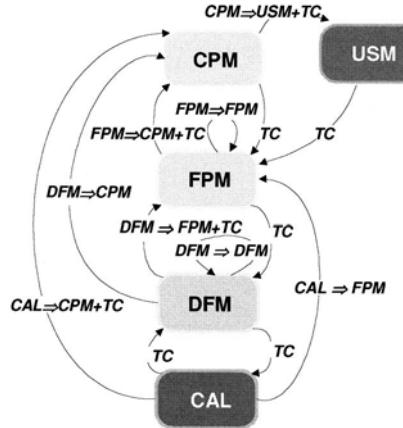


Figure 2. Transition diagram of DFAC operating modes.

The paper will briefly describe DFAC architecture and algorithms which have been designed and tested in view of the successful Preliminary Design Review (PDR), March 2002. The achieved performance, presented in this paper, gives a positive perspective on the solidity of the current DFAC design. The first section explains the DFAC mode architecture and provides a description of the mission requirements. The second section deals with the critical and innovative GOCE control sensor and actuators: gradiometer and electrical thrusters. The last section is devoted to design methodology, algorithms and performance of the main mission mode, the Drag-Free Mode (DFM), which is specifically intended to enable gradiometer calibration (COP) and operation (MOP) during the 6-months measurement phases.

2. Control Mode Architecture and Requirements

Figure 2 reports the transition diagram of the DFAC operative modes. The acronym TC indicates a transition which can be enabled from either on-ground commands or on-board software commands.

Coarse pointing mode, CPM. At the separation from the launcher, the spacecraft angular rate must be damped to values close to the orbital rate ($\cong 1.2 \frac{\text{mrad}}{\text{s}}$). At the same time Sun must be acquired. A maximum duration is allotted to spacecraft attitude for converging to LORF such to avoid complete battery discharge and too fast orbital decay. Rate damping is handled through MaGnetoMeter (MGM) and Sun Sensor (SS) data. Sun acquisition, employing Earth Sensor (ES) data, can be performed as far as the body *Z* axis is locked to nadir and the body *X* axis is aligned with spacecraft velocity. Spacecraft attitude is actuated by Magnetic

TABLE I
Coarse pointing mode (CPM) requirements

Variable	Unit	Bound	Comments
Solar aspect angle	rad	<0.7	In whichever season
Maximum angular rate error	mrad/s	<1.8	Damping to LORF in less than 15000s
X axis alignment to velocity	rad	<0.2	
Autonomy	days	8	No link to ground
Time to detect failures	s	<900	After separation

TABLE II
Fine pointing mode (FPM) requirements

		Preparation to DFM Orbit transfer maneouvres and hibernation	
Variable	Unit	Bound	Bound
Angular acceleration	$\frac{\mu\text{rad}}{\text{s}^2}$	<1	NA (Not applicable)
Angular rate	$\frac{\text{mrad}}{\text{s}}$	<0.87	<0.87
Attitude	mrad	<18	<1.8
Time to converge	s	<5400	One orbital period

Torquers (MTR) without relying on magnetic field models to assure the greatest autonomy. CPM and USM requirements are summarized in Table I.

Fine pointing mode, FPM. Once attitude has become sufficiently quiet, the Satellite-to-Satellite Tracking Instrument (SSTI), actually a GPS receiver, and the Star Tracker (STR) can be switched on and spacecraft attitude can be refined taking advantage of improved measurements and of full controllability permitted by Micro Thruster Assembly (MTA). During this mode, the Ion Thruster Assembly (ITA) will be commanded by ground to a constant thrust level equal to the expected X axis drag in order to reduce orbit decay and avert gradiometer saturation. As soon as spacecraft will be locked to LORF in a finer way, in other words linear and angular accelerations and pointing errors will settle below FPM thresholds, gradiometer will be switched on without risk for its safety. The FPM has been designed as a modular mode capable of arranging different actuator and sensor combinations. Accordingly, it can work by employing either MTA or MTR alone or MTA and MTR together. Similarly, SSTI can be replaced by an on-board orbit propagator, initialized from ground, and STR can be replaced by a proper integration of SS-ES-MGM data, Sun propagator and Earth magnetic field model.

TABLE III
Drag-Free Mode (DFM) requirements

Variable	Unit	Axis	<5 mHz [unit]	MBW [$\frac{\text{unit}}{\sqrt{\text{Hz}}}$]	0.1÷5 Hz [$\frac{\text{unit}}{\sqrt{\text{Hz}}}$]	5÷100 Hz [$\frac{\text{unit}}{\sqrt{\text{Hz}}}$]	Max value [unit]
COM acceleration	$\frac{\mu\text{m}}{\text{s}^2}$	X, Y	0.5	0.025	0.2	0.05	1
Angular acc.	$\frac{\mu\text{m}}{\text{s}^2}$	Z	0.5	0.025	0.2	0.05	0.5
Angular rate	$\frac{\mu\text{rad}}{\text{s}}$	X, Y, Z	1	0.015	0.025	0.005	1
Attitude in LORF	μrad	X, Y, Z	10	1			
LORF in inertial frame	μrad	Y	1200	0.5			
		X, Y, Z	370	10			
		X, Y, Z	200	15			

Drag-free mode, DFM. The mode, which is intended for scientific data collection, fully exploits gradiometer measurements to clean orbital motion from non gravitational accelerations and spacecraft attitude from all perturbing accelerations due to thermosphere drag, gravity gradient and magnetic field. The basic measurement set is represented by 18 linear accelerations coming from the 6 proof masses of the gradiometer, namely three measures for each mass along the gradiometer reference frame which can be taken as aligned to spacecraft body frame. Common mode and differential measures from different proof masses provide the spacecraft linear and angular acceleration vector. Gradiometer alone would not be sufficient for attitude control due the integrative effects of gradiometer drift and bias. To this end, attitude control loop must rely on STR and SSTI measurements. Thrusters represent the most critical DFAC equipments in the science phase, as outlined below. A single ion thruster is devoted to drag compensation along flight direction: drag is still quite large at those altitudes. Eight smaller electric thrusters are used to control the other five axes. Their commands are optimally distributed such as to minimize consumption and thrust level.

Calibration mode, CAL. To correct systematic gradiometer errors, scale, rotation and coupling factors, a calibration phase is planned before starting scientific data collection, at least at the first entrance of each Mission Operative Phase and every time gradiometer looks affected by large errors degrading the above factors. Calibration mode is exactly equal to DFM except the capability of shaking the satellite COM in a sinusoidal fashion at a fixed frequency (0.05 Hz).

Ultimate safe mode, USM. Control architecture is completed by the Ultimate Safe Mode which has been imposed for handling CPM failures and still guaran-

teeing sufficient power supply. In this mode, ES measurements are assumed not to be available as well as EMF and ephemerides models, causing the body X axis to form a quite large angle with respect to spacecraft velocity, the magnitude depending on drag intensity. Nevertheless, SS measurements allow to still keep the spacecraft Sun-oriented for battery power supply. Fall-back from each mode to CPM is triggered by attitude or angular rate anomaly detection. FPM behaves like a higher-level fall-back mode from DFM and CAL in case of anomalies not jeopardizing satellite safety. In fact, reacquisition of drag-free conditions is simpler starting from FPM, while climbing up from CPM could require a new calibration phase. Exit from USM is possible using the extended capability of the USM of working also without STR and SSTI. Ground-based initialisation is guaranteed by triggering mode exit only from ground.

3. Critical and Innovative GOCE Sensors and Actuators

The gravity gradiometer, developed by ONERA, France, is an ensemble of 6 three-axis electrostatic accelerometers. In each accelerometer a 320 g proof mass is electro-statically suspended and actively controlled (at 1 kHz rate), in 6 degrees-of-freedom, at the centre of a cage, by means of voltages applied to 8 pairs of electrodes machined on the internal walls of the cage. A thin (5 mm) gold wire, fixing the proof-mass electrical potential, is the only mechanical link. Capacitive sensors measure the proof mass displacement relative to the electrodes. The control voltages are proportional to the mass accelerations relative to the cage within the closed-loop control bandwidth of about 20 Hz. The $3 \times 6 = 18$ restoring accelerations of the 6 proof masses constitute the raw measurements, which after suitable filtering, are dispatched to a pair of different channels: the scientific channel sampled at 1 Hz rate and the control channel sampled at 10 Hz rate. This approach reduces cost and masses, but the higher sampling rate of the control channel imposes stringent requirements to instrument noise sources, as for instance to command DAC (digital-to-analog conversion) noise, at the limit of space qualified components. High thermal stability and low electro-magnetic susceptibility are the means for noise control.

Thruster selection for DFAC and orbit maintenance has been one of the major design driver of the platform and subject of a lengthy and complex trade-off. Thruster assembly requirements are:

- extended thrust range: 1 to 20 mN and more along X linear axis,
- high resolution: below $2 \mu\text{N}$ for Y and Z linear axes and rotation axes,
- fast response: up to 20 Hz along X linear axis,
- low noise and continuous operation,
- high efficiency with respect to required mass, propulsion system and consumable, and power.

Electric propulsion has been kept as the best solution to meet such requirements, but a few problems are still unsolved and subject to investigation. Thrust modulation for controlling the linear X acceleration (along the flight direction) and for orbit raising manoeuvres is achieved through the controlled emission of a Xenon ion-beam from a single equipment. Similar ion thrusters have already flown as for instance on Deep Space 1 (NASA) and, more recently, on Artemis (ESA). The innovative GOCE issue is represented by the large dynamic range required to compensate thermosphere drag at the operative mission epoch and altitude. Critical issues are instead grid erosion and beam-out. The latter effect corresponds to ion flow interruption, caused by internal short-circuits due to accumulation of residual particles on the ion accelerating grids as well as to electrostatic discharge between accelerating grids and screen. A mean time between consecutive beam-outs of about 15 to 20 minutes has been experienced on Artemis. Such figures being not compliant with GOCE mission demanding smooth and quiet acceleration environment, investigation is in progress to find appropriate solutions.

Micro-thrusters relying on FEEP (Field Emission Electric Propulsion) technology have been recently selected for the drag-free control along Y and Z linear axes and for attitude control. The Phase A baseline, cold gas propulsion, was abandoned during the last design phase because of high propulsion noise (mainly related to moving masses of the propulsion components) in contrast with DFM requirements. Two main families of FEEP thrusters are nowadays available depending on their consumable, Cesium or Indium. Cesium permits larger thrust levels, which is a key point for managing uncertainties about environmental disturbances as imposed by possible mission delays, but it is quite sensible to atomic oxygen and ground environment humidity. Indium permits better thrust resolution independent on the thrust level. The need for a precise measurement of their low thrust levels, in the order of μN , and of their noise profile has led to the design of a new test-bench, the Nanobalance, under test at the Metrology Institute ‘G. Colonnetti’, Turin, Italy.

4. Drag Free Mode

4.1. DESIGN AND ARCHITECTURE

DFM architecture and algorithms have been designed through a state-equation method capable of harmonizing analytic and simulation design (Canuto, 2002). The central point is the construction of two coherent dynamic models of the spacecraft to be controlled: the fine model, acting as a surrogate of the unavailable spacecraft, and the design model. The design model is a set of discrete-time state equations, which are at the same time the core of the control algorithms and provide the mathematical framework for analytic design. The unmodelled dynamics, together with their assumed uncertainty, is defined as the discrepancy between fine and design model. The key guideline is that the uncertain dynamics, filtered by

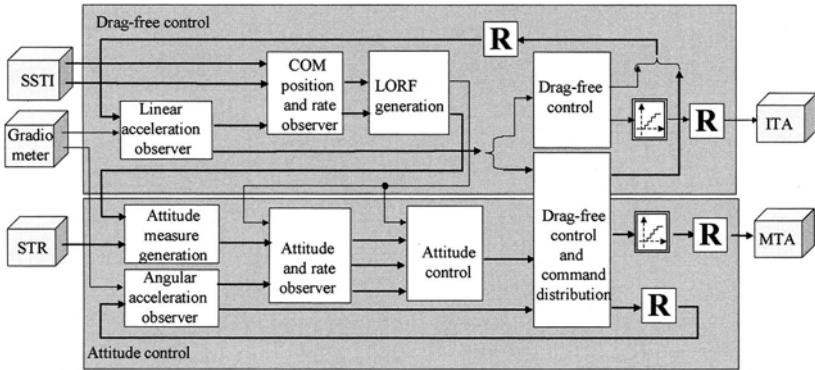


Figure 3. High-level block diagram of Drag-Free Mode (DFM) control algorithms.

a suitable state observer can be made to approach the controllable dynamics of the design model within a predefined tolerance thus making applicable and robust state space control strategies. Thrusters and sensors (specifically the gradiometer) are the main sources of unmodelled dynamics. The conceptual design steps are.

- Construction of the design model, including controllable and disturbance dynamics for a total of 42 state equations, definition of the performance variables and of the reference trajectories of the controllable state variables.
- Design of the reference trajectory generator providing the local orbital reference frame (LORF) and of the attitude and drag-free control algorithms in the form of state feedback loops on the design model. Optimization of the command distribution to ITA and MTA, such as to minimize average and peak thrust level.
- Design of the state observers and tuning of their eigenvalues for guaranteeing 'spacecraft' closed-loop stability against the unmodelled dynamics.
- Performance and robustness tests on the simulated 'fine' model.

A standard architecture, respecting the above conceptual steps, has been pursued (see Figure 3). Gradiometer measures are sampled at the highest rate, 10 Hz. Micro-thruster commands are computed at the highest rate and then dispatched to ITA and MTA at their own rate, 10 Hz and 2 Hz respectively. SSTI and STR measurements, sampled at a lower rate, 1 Hz and 2 Hz respectively, are treated according to the usual predictor-corrector scheme. Predictor works at the highest rate, corrector at the measurement rate.

Two decomposition patterns have been adopted to deal with model complexity:

- *Hierarchical decomposition*. The drag-free control is a wide-band inner feedback loop (gradiometer-to-thruster), which can be designed separately from attitude control. Attitude and rate control is instead a narrow-band outer feedback loop (STR to thrusters) to compensate slow varying drifts, like gradiometer ones, below certain threshold frequencies, roughly approaching the lower extreme of the MBW, 5 mHz.

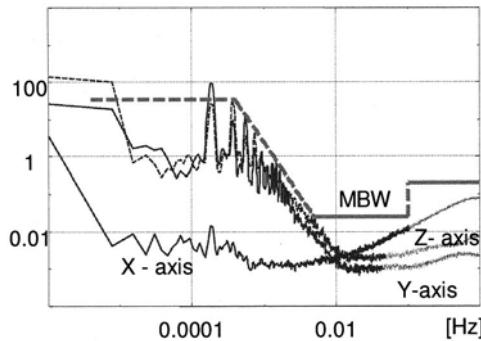


Figure 4. Power spectral densities (PSD) of the residual linear accelerations compared to target bound. Ordinates: $[\frac{\mu\text{m}}{\text{s}^2\sqrt{\text{Hz}}}]$. Abscissae: Frequency [Hz].

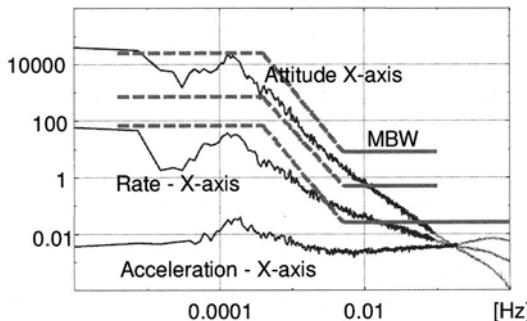


Figure 5. Power spectral densities (PSD) of the roll attitude $[\mu\text{rad}]$, of the relevant angular rate $[\frac{\mu\text{rad}}{\text{s}}]$ and residual angular acceleration $[\frac{\mu\text{rad}}{\text{s}^2}]$. Ordinates: $[\frac{\text{unit}}{\sqrt{\text{Hz}}}]$. Abscissae: Frequency [Hz].

- *Coordinate decomposition of the attitude control.* Inner and outer feedback loops can be decomposed into three parallel loops, one for each attitude coordinate. Cross-couplings have been treated as known internal disturbances.

4.2. PDR PERFORMANCE

Two main issues were the concern of PDR. First was robustness of DFM accuracy performances in front of spacecraft and environment uncertainties, like mass and inertia variations, thrust misalignments, thermosphere density variation due to variable launch epoch and MOP altitude. Figures 4 and 5 show the unilateral root power spectral densities (PSD) of the residual linear accelerations and of the different attitude variables, angular accelerations, rates and attitude angles, under worst-case spacecraft and environment conditions. They are compared to target bound profiles, derived from Table II.

Target accuracy is met with a good margin in the most significant region of gradiometer measurements, i.e. from MBW and to control Nyquist frequency, 5 Hz.

TABLE IV
Comparison of command distribution strategies

Thrust minimization strategies	ITA [mN]		MTA [mN]	
	Average	Peak	Average	Peak
None	9.78	20.55	2.29	1.06
Peak			1.26	0.55
Peak and average			0.93	0.51

Accuracy appears somewhat critical at low frequencies for what concerns cross-axis residual accelerations. But this is the price to pay to thrust-peak reduction strategy. The DC component of the attitude PSD is due to worst-case STR bias.

Second were compatibility and sizing of electrical micro-thruster technology to allow the requested thrust authority with some margin. Table IV shows the progressive reduction of the average and peak thrust requested to MTA, by applying different optimal command distribution strategies.

5. Conclusions

DFAC has been integrated into the end-to-end GOCE simulator, providing the above performances. Tests have shown that accuracy performances are robust against spacecraft and environment uncertainties, and command distribution strategies are very effective in thrust level reduction.

Acknowledgements

The authors would like to mention Massimo Palomba and Massimiliano Saponara of Alenia Spazio, Basilio Bona, Giuseppe Calafiore and Marina Indri of Politecnico di Torino for their valuable work in the definition of DFAC algorithms. A particular thank also to the Alenia DFAC and propulsion system engineers Antonio Tramutola and Mario Pessana for their precious suggestions.

References

- Canuto, E.: 2002, 'Sub-nanometric optics stabilization in view of the GAIA astrometric mission'. *Control Engineering Practice*.
- Johannessen, J. A, M. Aguirre-Martinez: 1999, 'Gravity Field and Steady-State Ocean Circulation Mission'. In: *The Four Candidate Earth Explorer Core Missions*, ESA SP-1233, July 1999.

ON SUPERCONDUCTIVE GRAVITY GRADIOMETRY IN SPACE

STEFAN ZAREMBIŃSKI (f4asz@fy.chalmers.se)

Department of Physics, Chalmers University of Technology, S-412 96 Gothenburg, Sweden

Received: 15 May 2002; Accepted in final form: 23 December 2002

Abstract. Superconductive gravity gradiometers for space have been expected in vain for more than fifteen years, since the niobium superconductor technology has matured and became commonly regarded as adequate to the purpose. Therefore an old design idea may be suspected of inefficiency, and new concepts should be considered. We propose one that abandons an attractive yet restrictive method of spatial differentiation by the principle of the magnetic flux conservation. Instead, it uses the SQUID just as a low noise sensor in the test mass displacement transducers, and differentiates by means of a negative feedback. We argue that the feedback can ease the known obstacles. Especially, it provides convenient practical means for neutralization of the low frequency SQUID noise, and for correction of dimensional inaccuracies of the sensor mechanics. Moreover, the feedback can organize a cluster of twelve elementary accelerometers into a precise tensor gradiometer that can self-correct its inaccuracies by tuning the cross-coupling between its elements.

1. Introduction

Superconductive gravity gradiometers can incorporate a SQUID (Superconducting Quantum Interference Device) as its innermost transducer. The SQUID is a low noise current sensor. Superconductivity in space can benefit from high temperature ceramic superconductors (HTS) which feature transition temperatures as high as 90 K, in contrast to the low temperature metallic superconducting (LTS) materials that are operational at about 4 K.

Superconductivity greatly improves the chance of a successful design of a sensitive gravity gradiometer. An essential advantage is that the nearly lossless superconductive circuitry of the test mass displacement transducers does not contribute to the thermal noise of the instrument. With a better sensitivity of the transducer, a gravity gradiometer can be lighter and smaller. Moreover, the amount of heat dissipated by a wholly superconducting gravity gradient sensor with SQUIDs can be extremely low, thus limiting thermal flows in a small working chamber of an already compact mechanism and, as consequence, improving conditions for thermal stability of the instrument, which is the most limiting factor finally.

Superconductive gravity gradiometry in space is not practical at present but it is commonly seen as a technologically feasible, natural development of the common techniques. This view is based on reliable analyses. According to one (Zarembiński, 2001), calculated thermal noise of an instrument made by means of the present HTS technology, with the test masses of 0.1 kg, suspension resonant fre-



quency of 0.1 Hz, and with the base length of 0.5 m, corresponds to a gravity gradient signal of about $10^{-26} \text{ s}^{-4}/\text{Hz}$. It is obviously impossible to reach absolute limits but even with a sensitivity 10^4 times worse than the limiting value, and with 10 times larger test mass of 1 kg, it is enough to detect a 10 km cube of ice from a 300 km orbit.

Superconductive gravity gradiometers for space have been sought for twenty years, always seen just within researchers' reach, but still they are as impractical as they were fifteen years ago. We want to suggest that the evident impasse can be overcome only with new ideas for both the instrument construction and the logistics of the problem, which is a sum of human and material resources, planning and coordination sufficient to lead this complex problem to its solution. However, here we can discuss the construction issue only.

For this discussion we divide the question in two parts. One part is the technology background of the problem. In our opinion, the present technology is essentially sufficient. It only needs a refinement, by means of the logistics rather than by research breakthroughs. We point out briefly the areas of concern in Section 2.

The other part of the problem is an efficient concept that could develop a practical instrument from the available technology. In Section 3, we consider a design idea that could organize a superconductive gravity gradiometer. This scheme solves the known deficiencies of the best conventional designs, and also combines their strongest features.

2. Technological Background

Space cryogenics for SQUID instruments is equally primitive in both LTS and HTS temperature domains. Only open systems, irreversibly vaporizing a load of a coolant are practical. Therefore, a SQUID gravity gradiometer for space is commonly planned as a huge tank. A heavy cooling unit can reduce the performance of a superconducting gradiometer below a level achievable for ambient temperature instruments of the same total mass.

It is hard to believe that a closed-cycle refrigerator, pumping about 0.1 W between 60 K and 150 K, can be a great technical problem nowadays, but apparently, it is such. A difficulty is in a need for ultra quiet pumping action of the machinery, without vibrations. We believe that a prompt solution of this problem can be achieved, if a properly organized action is undertaken. Even if it fails, still a less than perfect refrigerator can be used, being run periodically between measurement phases. With a closed-cycle cooler, a sensitive HTS instrument can be light and compact, and long-lived enough to go to planetary missions.

The LTS technology is mature. It is accessible on commercial basis, reproducible and reliable. On the contrary, a space-quality HTS SQUID sensor for a sensitive gravity gradiometer is impossible presently. The present HTS technology is efficient, yet it is unstable. This efficiency is distributed among scattered

laboratories, each one can do an impressive fragment sometimes, but reproducibility is never granted. Brilliant entities are reported occasionally (Schilling *et al.*, 1996), however, even they discontinue. The problem of the technique that was never refined by the industry is in methods that it uses which are of a provisional character: very slow and inefficient. Therefore it is practically impossible to finish a complex task, involving many process steps for each single test, and a multi-test optimization procedure, while such is to be the work on an HTS SQUID for a space gravity gradiometer.

In practical conditions, the sensitivity (Zarembiński, 2001) of a SQUID gravity gradiometer is inversely proportional to the magnetic flux $\Phi = L_s I_0$ through the sensing loop of the inductance L_s , with a current I_0 . The technology and physical properties of a superconductor impose limitations on the current density, therefore the current I_0 is strictly limited, and hence the inductance L_s cannot be small. However, there is an essential conflict between the magnitude of the sensing loop and efficiency of the flux transformer, which links the SQUID sensing and interferometer loops. This efficiency rapidly falls with an increase of the working temperature. For 60 K, the transformer can be quite complicated (Zarembiński and Claeson, 1992), needing an extensive experimental optimizing process. Therefore the only way towards an efficient HTS SQUID starts now with an upgrading of the present technological tools up to a quasi-industrial level.

3. Organizing Idea

A gravity gradiometer in space should be six-dimensional, equally sensitive to all components of the gravity gradient tensor. Such an instrument not only provides complete information about the gravity field, but it can also separate an effect from unavoidable uninertiality of the satellite movement. Otherwise, with an incomplete gradiometer, a spacecraft has to be equipped with an auxiliary instrumentation able to provide data missing in the gradiometer output.

In all reported studies, the gradiometer is considered as an aggregation of separate dipoles, each sensitive to a particular component of the gradient, and therefore the main effort is concentrated there on a single dipole, in hope that a complete instrument can be combined later. We will argue that treating a tensor instrument as an entity can help to solve difficult problems specific to an isolate gradiometric dipole.

A gravity gradiometer in space needs to be sensitive to a low frequency signal of about 10^{-3} Hz. Unlike the capacitive transducers (Bernard and Touboul, 1991) with high frequency capacitive-inductive bridges that convert the signal up to the carrier frequency, the conventional superconductive differentiating transducers (Chan and Paik, 1987a) mix the gradiometric signal with low frequency ($1/f$) noise of their SQUIDs. Low temperature SQUIDs have $1/f$ knee at about 0.1 Hz, the high temperature knee is at 1 Hz, in both cases much above the lower signal frequency band

limit. We will introduce a differentiation method that consists with an $1/f$ SQUID noise suppression scheme.

In a quest for a practical HTS gravity gradiometer, the conventional low temperature design tradition is useless. The two worlds differ in many aspects, but one point of difference is enough here. The classic layout essentially relies on extended wire superconducting circuits, with large sensing wire coils, while the HTS electronics is practical only in thin film structures, even if HTS cables are sometimes reported. We will suggest an HTS appropriate design.

Sometimes it happens that a lucky concept can reduce a host of essential design problems. Such an idea can be called the *organizing design idea*. We want to suggest that a strong feedback, interconnecting a cluster of twelve accelerometers can be an organizing idea for a superconductive gravity gradiometer, proper for either low or high temperature applications. To introduce it, we start with a brief outline of the tradition.

3.1. GRAVITY GRADIOMETER DIPOLE

A gradiometric dipole has two test masses suspended elastically from the instrument frame. It is essential that the suspension suppresses the freedom of the mass movement, ideally down to differential movement, as it is in the case of the torsion balance, shown in Fig. 1(a). There the differential rotation of a pair of dumbbells

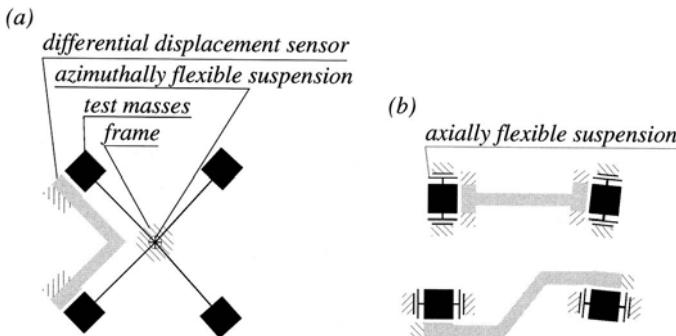


Figure 1. Configurations of a gravity gradiometer dipole. (a) Torsion balance pair, sensitive to an *off-diagonal* component of the gravity gradient tensor. The orthogonal dumbbells can swing separately, and therefore the output of a differential sensor is independent of the rotational acceleration. (b) The test masses are separately suspended from the instrument frame by axially elastic, cantilever structures. The first dipole is sensitive to a *diagonal* component of the gravity gradient tensor, while the second one is an *off-diagonal* sensor. The sketches are distorted asymmetrically, to represent manufacturing errors.

is proportional to an (*off-diagonal*) component of the gravity gradient sensor. In part (b) of the figure, the test masses are separately suspended from the frame by axially elastic cantilever structures, and therefore they are confined to the linear movement. The differential move of the masses can be a measure of a gravity gradient component, corresponding to the orientation of the suspension axes. Both

dipole schemes, (a) and (b), have practical representations, implementing low temperature SQUIDs. The in-line form from Fig. 1(b) has been devised and used in an experiment by Chan and Paik (Chan and Paik, 1987b). The torsion balance form is being successfully developed (Kann *et al.*, 1980) as an airborne tool for geological prospecting.

Both practical superconducting gravity gradiometers use differentiating SQUID transducers of the same sort (Chan and Paik, 1987a), represented schematically in Fig. 2. The transducers have SQUIDs as the null sensors in superconducting bridges of inductors, ingeniously arranged so that they are sensitive only to the differential move of the test masses.

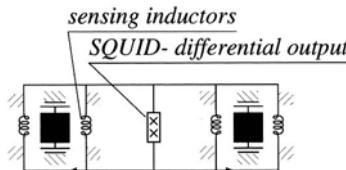


Figure 2. Differentiating SQUID transducer. The test masses have superconducting sides, interacting with permanent currents of opposite orientation in the sensing circuits. If the test masses move commonly, as the result of the background acceleration, then the current through the SQUID does not vary. It does, if the moves differ. The variation is a measure of the gravity gradient.

3.1.1. Limitations and Potential of Original Designs

Table I collects characteristics of two real instruments, and confronts them with expectations specified by a space agency (Roussel, 1993), and with physical limitations resulting from the thermal noise of the SQUID. The practical results are much below the demands of the space quality, which are close to the absolute limit set, for the given instrument dimensions, by the SQUID thermal noise (Zarembiński, 2001). The thermal noise PSD is proportional to the SQUID junction equivalent energy sensitivity and to the squared test mass suspension natural frequency, and it is inversely proportional to the test mass and squared base length.

Both related designs share the same vulnerability to the $1/f$ SQUID noise. Therefore, the deterioration of the sensitivity, documented in table I, can be ascribed to this noise contribution.

The CMRR factor of the Maryland (Chan and Paik, 1987b) instrument is too low. This is evidently the result of a lack of a correction system that could counterbalance the misalignment of the test mass suspensions. A variety of dimensional errors is shown in Fig. 1. The differentiating SQUID transducer in Fig. 2 equalizes the suspension resonant frequencies. Misalignment of the suspensions in the Maryland model, or misplacement of the mass center in the Western Australia device needs additional regulation systems.

The superiority of the torsion balance over the instruments using linear suspensions results from the ease of having the torsion suspension very soft rotatory and, in the same time, transversely rigid. With its deciding advantage, the torsion

TABLE I

Experimental characteristics of helium cooled gravity gradiometers, compared with expectations of a space agency, and with physical limitations resulting from the thermal noise of the SQUID. For the estimate, the SQUID junctions equivalent energy sensitivity is set to 8×10^{-33} J/Hz.

	Agency [†]	Maryland [‡]	W.Australia [§]
test mass (kg)	—	0.40	0.34
natural frequency (Hz)	—	19	3
base length (m)	—	0.16	0.045
sensitivity ($\text{s}^{-2}/\sqrt{\text{Hz}}$)			
at 1 Hz	10^{-12}	7×10^{-10}	5×10^{-10}
at 0.1 Hz	10^{-12}	7×10^{-10}	5×10^{-10}
at 0.01 Hz	10^{-12}	3×10^{-9}	10^{-9}
at 0.001 Hz	10^{-12}	$10^{-8}(\star)$	10^{-8}
SQUID limit [§] ($\text{s}^{-2}/\sqrt{\text{Hz}}$)	—	1.3×10^{-13}	8×10^{-14}
CMRR	10^6	$3 \times 10^3 - 3 \times 10^4$	10^9

[†] (Roussel, 1993); [‡] (Chan and Paik, 1987b); [§] (Kann *et al.*, 1980);

[§] (Zarembiński, 2001); [★] extrapolated.

balance layout still has to compete with the linear one. It is so, perhaps, because the torsion balance is incorrectly conceived as an *off-diagonal* sensor. We will show that a complete tensor gradiometer can be set of six torsion balance dipoles.

3.2. SYNTHESIZED GRADIOMETRIC DIPOLE

A gradiometric dipole can be synthesized of two accelerometers by means of a strong negative feedback, as it is shown in Fig. 3. One feedback loop releases the background acceleration. This feedback is controlled by a test mass displacement sensor, and produces a feedback force, balancing the acceleration force that tends to move the test mass.

The same feedback force is also applied to the other test mass of the dipole, but the mass is kept at its neutral point by the second, differentiating feedback loop. Hence the reaction of the differential feedback is proportional to the gravity gradient.

We consider electrostatic drives as feedback actuators, hence the differentiation accuracy results from the identity of the voltages powering the drives, if the mechanism of the instrument is balanced and stable (thermally, dimensionally, etc.). Under the same stability condition, in the original setup with the superconducting differentiating inductive bridge, the accuracy relies on the constancy of the magnetic fluxes frozen in the superconducting loops.

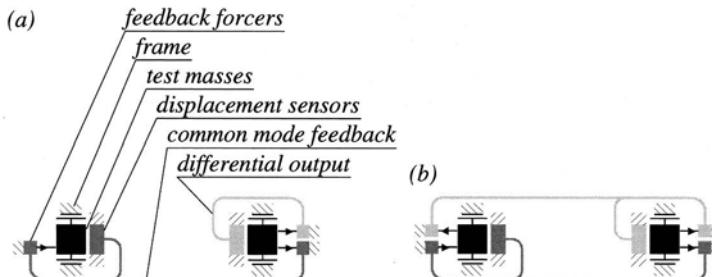


Figure 3. (a) The dark-shaded feedback loop is controlled by the left test mass displacement sensor, and acts in-phase on both masses equally. It releases the background acceleration. Then the light-shaded feedback corrects the differential displacement. Its output is proportional to the gravity gradient. (b) A symmetrized version of the gradiometer in panel (a). The differential feedback acts off-phase on both test masses.

The traditional operating principle, forcing a need for the detection of low frequency current changes in the SQUID interferometer, causes the problem of an interference between the $1/f$ SQUID noise and the gravity gradient signal. With the synthesized dipole layout, it is possible to use the lock-in method, with an alternating current in the sensing inductive bridges. Thus the signal is converted up above the noise, passed through the SQUID, and subsequently shifted back, separated from the noise.

Unlike it is in the original setup, there is no extended superconducting link between the test masses in the synthesized dipole version. Therefore, the superconducting electronics can be integrated in reliable thin film structures.

In the original design, gradiometric dipoles are balanced by the adjusting of the ratio between the superconducting currents in the sensing inductors. The same method could be used within the synthesized dipole layout, even if it can be impractical with thin film SQUID transducers. Alternatively, a gravity gradiometer with feedback can be balanced by the scaling of its feedback actuators. With electrostatic drives this scaling can be made by the positioning of a small fragment of the frame-side electrodes, with a piezoelectric element, possibly. One more *virtual* balancing method, proper for a feedback gravity gradiometer, is explained in Section 3.3.

3.2.1. Torsion Balance Dipole

The torsion balance configuration has exceptional design advantages. It has the inherent dipole symmetry, and the torsion suspension can separate very strictly the modes of the test mass motion. A pair of torsion balances can measure two components of the gravity gradient tensor. Especially, if the angle between the dumbbells is 90° , the case is shown in Fig. 1(a), the sensitivity is degenerated to an off-diagonal component, but if the angle is $\arctan(2) \approx 63^\circ$ then the system is equally sensitive to both planar components.

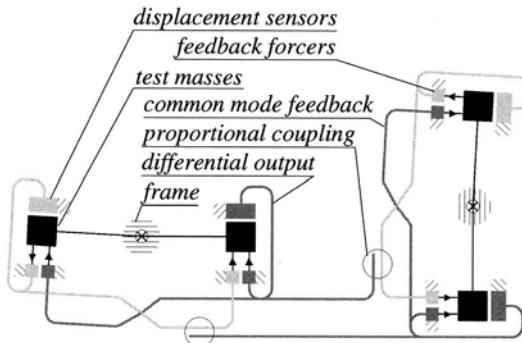


Figure 4. Correction of a misalignment error in a synthesized gradiometric dipole. The effect of the error is weakly proportional to the orthogonal component of the background acceleration, which is measured in the common mode feedback loop of the corresponding sensor, and coupled off-phase to the differential feedback actuators of the dipole.

3.3. TENSOR GRADIOMETER

Dimensional inaccuracy of a gradiometric dipole results in a false gradiometric signal, proportional to the background acceleration. Therefore, the inaccuracy error can be corrected by means of a calibrated coupling between the gradiometer and a background acceleration sensor. A synthesized gravity dipole can be used as an accelerometer, because it measures a component of the background in its common mode feedback loop. A set of three synthesized dipoles provides all the information needed for the correction of the dimensional inaccuracy of a tensor gradiometer.

Figure 4 shows a simplified case of the misalignment error. The false signal output from the differential feedback loop of a dipole is proportional to the orthogonal component of the background, which is measured in the common feedback loop of the corresponding sensor, and coupled to the differential feedback actuators of the dipole.

The accuracy of the correction has to be about the CMRR reciprocal which is 10^{-6} (Roussel, 1993). If the precision of preadjustment is 10^{-3} , then the accuracy of the correction coupling has to be also 10^{-3} . With electrostatic feedback drives, the coupling ratios can be set by potentiometers with an accuracy much better than necessary.

A gravity gradiometer can be conveniently tuned by use of a set of shakers forcing the instrument to vibrate. The vibrations simulate the effect of the background acceleration, and the instrument is adjusted to minimize its response to them. An instrument with synthesized dipoles can use each of its feedback actuators for shaking itself.

A complete tensor gradiometer needs twelve test masses, a pair for each tensor component, or a set of three pairs of torsion balances. Such a set can be neatly assembled, as it is schematically shown in Fig. 5(a). Panel (b) of the same figure presents a basic design idea of the gradiometer mechanics, proving that the cluster

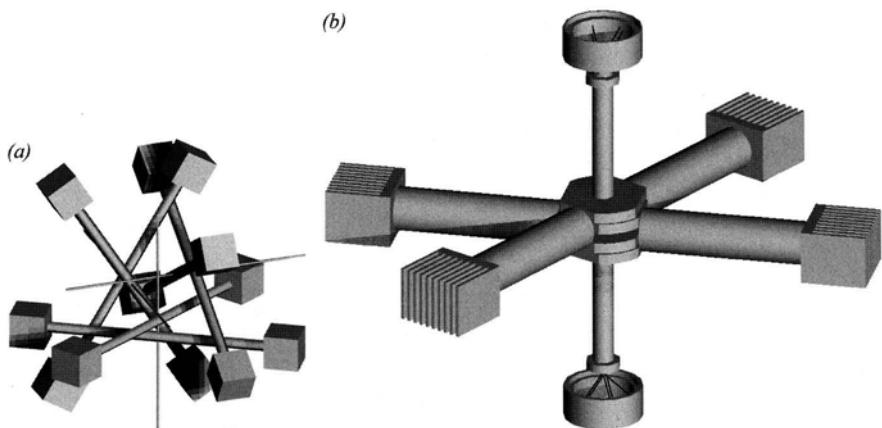


Figure 5. (a) A cluster of three pairs of torsion balances makes a complete tensor gravity gradiometer. (b) A design idea of the tensor gravity gradiometer mechanism. The test masses are segmented, as they are electrodes of the feedback actuators. The complementary electrodes, not shown, are mounted on the frame. The masses are rigidly assembled in pairs, mounted on structures that interlace each other. The structures have tubular coaxial shafts in their middles, extended up and down, each one having its thinner counterpart led inside. The shafts are torsionally suspended from the instrument frame by thin strings, arranged as tripod legs at the ends of the shafts. The suspension strings are anchored in the instrument frame, of which only small elements are in the figure. Each dumbbell of a pair can independently swing around the common axis, in accordance with the scheme in panel (a) of the figure. The diameter of the whole structure is about 40cm.

configuration is practical. There is shown the suspension of a pair of the torsion balances. The suspension bearings belong to the instrument frame that is not in the picture. The test masses are segmented, formed as electrodes of the feedback actuators. The electrodes have their complementary counterparts mounted on the frame, together with the test mass displacement transducers.

3.3.1. Potential of the Organizing Idea

The negative feedback allows to shift the complexity of a design from the mechanics to the computer software. The mechanics of a synthesized tensor gradiometer can be most simple and highly modular. All the complicated interaction between the hardware modules occurs inside a computer that closes the system feedback loops.

The feedback virtually freezes the dynamics of the gradiometer, counterbalancing external forces. A static mechanism can be simplified. For example, parasitic interaction between parts, such as crosstalk between test mass position sensors, can be easier reduced to a safe level, the test mass suspensions can be planned without great care about linearity.

The negative feedback can linearize the controlled system, and can make it dependent only on the *feedback filter* which can be a strictly confined part. The

stronger the feedback is, the more accurate is the linearization. The maximal strength of the feedback depends on the speed of the feedback loop.

In a synthesized gradiometer, the sensor mechanism itself is the feedback filter. Hence, if the feedback loop is strong enough then the quality of a gradiometer depends solely on the stability of the mechanics which is a crucial factor in any design, in any case. Then ultimately the sensitivity of a gravity gradiometer is limited by the stability of the working temperature. With superconducting electronics that produces a very low amount of heat, the instrument working chamber can approach the adiabatic state. We have learned from a comprehensive simulation of an HTS feedback synthesized gravity gradiometer that the working temperature can be kept within a margin of $5 \mu\text{K}$ at 60 K. In this condition, with 0.4 kg test mass and 0.4 m base length, a sensitivity of $10^{-24} \text{ s}^{-4}/\text{Hz}$ can be expected.

The idea of the feedback synthesized gradiometer is independent of the nature of the innermost transducers considered for the instrument. It is valid for superconducting SQUID sensors, either LTS or HTS, and for capacitive sensors with field effect transistor preamplifiers, as well.

References

- Bernard, A. and Touboul, P.: 1991, 'The GRADIO accelerometer: design and development status', In: C. Mattoe (ed.): *The Solid-Earth mission Aristoteles*, ESA Publications Division Nordwijk, Netherlands, ESTEC, 61–67.
- Chan, A. A. and Paik, H. J.: 1987a, 'Superconducting gravity gradiometer for sensitive gravity measurements. I. Theory', *Phys. Rev. D* **35**, 3551–3571.
- Chan, A. A. and Paik, H. J.: 1987b, 'Superconducting gravity gradiometer for sensitive gravity measurements. II. Experiment'. *Phys. Rev. D* **35**, pp. 3572–3597.
- Roussel, P.: 1993, 'Opening of workshop', In: J. Flokstra (ed.): *Proceedings of the workshop on high-T_c superconducting gravity gradiometers*, University of Twente, Twente, Netherlands, 7–12.
- Schilling, M., Krey S., and Scharnweber, R.: 1996, 'Biomagnetic measurements with an integrated YBa₂Cu₃O₇ magnetometer in a hand-held cryostat', *Appl. Phys. Lett.* **69**, 2749–2751.
- van Kann, F. J. et al.: 1990, 'Laboratory tests of a mobile superconducting gravity gradiometer', *Physica B* **165-166**, 93–94.
- Zarembiński, S. and Claeson, T.: 1992, 'Design of multi-loop input circuits for high T_c superconducting quantum interference magnetometers', *J. Appl. Phys.* **72**, 1918–1935.
- Zarembiński, S.: 2001, 'Sensitivity limitation of a superconducting quantum interferometer sensor-based gravity gradiometer', *IEEE Trans. Appl. Superconduct.* **11**, 4041–4047.

Address for Offprints: Stefan Zarembiński, Department of Physics, Chalmers University of Technology, S-412 96 Gothenburg, Sweden, f4asz@fy.chalmers.se

SATELLITE-SATELLITE LASER LINKS FOR FUTURE GRAVITY MISSIONS

P. L. BENDER (pbender@jila.colorado.edu), J. L. HALL

(jhall@jila.colorado.edu) and J. YE (ye@jila.colorado.edu)

*JILA, National Institute of Standards and Technology and University of Colorado, Boulder,
Colorado 80309-0440 USA*

W. M. KLIPSTEIN (klipstein@jpl.nasa.gov)

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA

Received: 15 May 2002; Accepted in final form: 15 August 2002

Abstract. A strong candidate for use in future missions to map time variations in the Earth's gravity field is laser heterodyne measurements between separate spacecraft. At the shortest wavelengths that can be measured in space, the main accuracy limitation for variations in the potential with latitude is expected to be the frequency stability of the laser. Thus the development of simple and reliable space-qualified lasers with high frequency stability appears to be an important goal for the near future.

In the last few years, quite high stability has been achieved by locking the second harmonic of a Nd:YAG laser to a resonant absorption line of iodine molecules in an absorption cell. Such a laser system can be made quite robust, and temperature related frequency shifts can be controlled at a low value. Recent results from laboratory systems are described. The Allan standard deviation for the beat between two such lasers was 2×10^{-14} at 10 s, and reached 7×10^{-15} at 600 s.

1. Introduction

The use of laser heterodyne measurements between separate spacecraft to map the Earth's gravity field and time variations in it has been discussed by a number of authors (see, e.g., Schumaker, 1990; Bender, 1992; Colombo and Chao, 1992; Watkins et al., 2000). As an example, the type of GRACE follow-on mission considered by Watkins et al. (2000) will be discussed briefly. To be specific, we assume the following: that the Allan standard deviation (fractional frequency fluctuation) for the laser frequency is 1×10^{-14} from 10 to 5000 s period; that the spurious acceleration noise level for the test mass in the accelerometer on each spacecraft is $10^{-12} \text{ m/s}^2/\sqrt{\text{Hz}}$ down to at least 0.001 Hz; that the spacecraft separation is 100 km; and that the altitude of the nearly polar orbit is about 300 km. The assumed Allan standard deviation corresponds to a spectral amplitude of the fractional laser frequency variations of

$$(\delta\nu)/(\nu) = (0.85 \times 10^{-14})(1\text{Hz}/f)^{0.5}/\sqrt{\text{Hz}}, \quad (1)$$

where f is the frequency of the variations.



Space Science Reviews **108:** 377–384, 2003.
© 2003 Kluwer Academic Publishers.

The lasers in the two spacecraft are assumed to be independently stabilized. The received beam from the other spacecraft is beat (heterodyned) against the local laser on each spacecraft, and the phase or frequency of the two resulting beat signals is measured as a function of time. Combining the results gives the biased range or integrated Doppler observable, and other observables can be obtained from this. The biased range is measured in terms of the laser wavelength, which provides the necessary scale factor. Thus fractional fluctuations in the laser frequency and wavelength translate directly into the same fractional fluctuations δL in the measured range L .

Given Eq. (1) and the 100 km spacecraft separation in our example,

$$\delta L = 0.85 \times 10^{-9} (1\text{Hz}/f)^{0.5} \text{ m}/\sqrt{\text{Hz}}. \quad (2)$$

The corresponding acceleration noise δa_L is obtained by multiplying by $(2\pi)^2$:

$$\delta a_L = 3.4 \times 10^{-8} (f/1\text{Hz})^{1.5} \text{ m/s}^2/\sqrt{\text{Hz}}. \quad (3)$$

This is larger than $2^{0.5}$ times the assumed test mass acceleration noise for frequencies above 1.2×10^{-3} Hz, so the laser frequency noise dominates the assumed spurious acceleration noise for higher frequencies.

The distance measurement noise due to the laser frequency noise can be written as

$$\delta L = 2.7 \times 10^{-9} (0.1\text{Hz}/f)^{0.5} \text{ m}/\sqrt{\text{Hz}}, \quad (4)$$

so it will be $2.7 \times 10^{-9} \text{ m}/\sqrt{\text{Hz}}$ or larger over the frequency range of 0.0002 to 0.1 Hz that is of interest for gravity field measurements in space. Since 2.5 or 3 times $10^{-9} \text{ m}/\sqrt{\text{Hz}}$ is about the satellite translational noise level δL_T often quoted as expected for the drag-free systems planned for several other missions, it appears that the satellite-to-satellite distance measurements probably can be made between the optical benches on which the accelerometers for the drag-free systems are mounted. Alternately, measurements could be made directly between the test masses in the accelerometers, as is planned for the LISA mission (LISA Study Team, 2000).

Another possible error source for the distance measurements is the attitude variations of the satellite. The laser wavefronts transmitted from satellite A will be observed at satellite B as nearly spherical waves centered on the transmitting satellite. However, the deviations from spherical waves due to the laser beam and optical element imperfections will cause variations in the apparent distance when the attitude of satellite A changes. If the rms phase error over the transmitting aperture is $\lambda/10$ and corresponds to pure astigmatism, and the transmitting aperture diameter is D, the rms size of the resulting distance measurement error δL_a at an angle θ to the axis is

$$\delta L_a = [1/8\sqrt{3}] [(\pi)^2] [\lambda/10] [(D/\lambda)^2] [(\theta)^2]. \quad (5)$$

A result almost the same as this has been given by Robertson et al. (1997). The numerical coefficient is somewhat different, but only because of a different assumption about the shape of the transmitted wavefront error.

For $\lambda = 1060$ nm, $D = 2$ cm, a static offset error in the pointing of θ_{dc} , and a pointing fluctuation $\delta\theta$,

$$\delta L_a = 54[\theta_{dc}][\delta\theta] \text{ m.} \quad (6)$$

If $\theta_{dc} < 2 \times 10^{-5}$ rad and we require $\delta L_a < 2.7 \times 10^{-9}$ m/ $\sqrt{\text{Hz}}$, then the requirement on $\delta\theta$ is that it be less than 3×10^{-6} rad/ $\sqrt{\text{Hz}}$.

The above pointing stability can be achieved fairly easily by using one of the approaches that have been developed for the LISA mission (Danzmann et al., 1998; Hammesfahr 2001). The laser beam from spacecraft A will be very bright at spacecraft B, and a portion of the received beam can be focused onto a CCD array with a fairly long effective focal length. Any change in the attitude of spacecraft B can be detected as a motion of the spot on the CCD array. The motion can be used at medium and low frequencies to correct the spacecraft attitude, while fast motions can be handled by piezoelectrically controlled tilts of the transmitted laser beam.

With the above laser frequency, spacecraft translation, and beam pointing errors, the total errors near 0.1 Hz would be about

$$\delta L = (5 \times 10^{-9})[(0.1\text{Hz}/f)^{0.5}] \text{ m}/\sqrt{\text{Hz}}, \quad (7)$$

and

$$\delta a = (2 \times 10^{-9})[(f/0.1\text{Hz})^{1.5}] \text{ m/s}^2/\sqrt{\text{Hz}}. \quad (8)$$

This “range acceleration” uncertainty is closely related to the gravity gradient uncertainty, for which the corresponding value is

$$\delta\Gamma = (2 \times 10^{-14})[(f/0.1\text{Hz})^{1.5}]/\text{s}^2/\sqrt{\text{Hz}}, \quad (9)$$

or

$$\delta\Gamma = (2 \times 10^{-5})[(f/0.1\text{Hz})^{1.5}] \text{ Eötvös}/\sqrt{\text{Hz}}. \quad (10)$$

2. Iodine Stabilized Laser

There are many types of laser frequency stabilization schemes that could potentially be used in a space mission. However, one particularly attractive candidate for use in the near future is a Nd:YAG laser operating at 1064 nm wavelength, with its second harmonic stabilized to a molecular transition in iodine vapor. The NPRO type of Nd:YAG laser with output power levels of up to 200 mW has been space-qualified. With such a laser operating at 100 mW, 1 or 2 mW of second harmonic

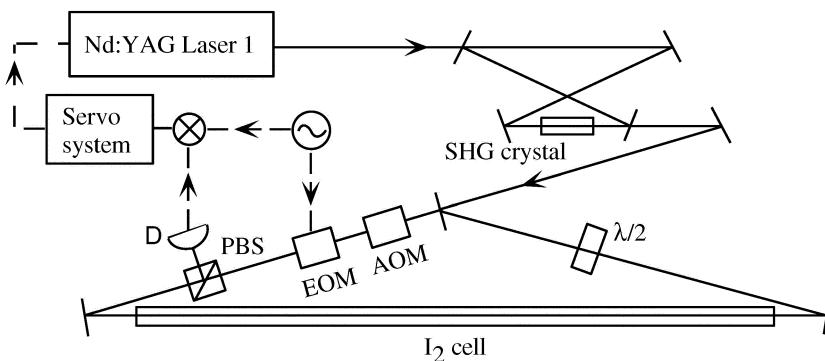


Figure 1. Diagram of a sub-Doppler iodine spectrometer: AOM, acousto-optic modulator; EOM, electro-optic modulator; PBS, polarization beam-splitter; D, detector; SHG, second-harmonic generation.

power can be generated with a periodically polled doubling crystal without requiring a power build-up cavity. This power level at 532 nm is sufficient for saturated absorption measurements in iodine vapor.

Measurements of the above type have been reported by Ye et al. (1999, 2001). Figure 1 shows a simple schematic of a sub-Doppler iodine spectrometer used for laser frequency stabilization. A saturating beam is sent in one direction through the absorption cell, and this causes a dip in the population difference between the lower and upper levels for the desired molecular transition. The width of the dip is determined by a combination of the natural linewidth, the collisional linewidth, the transit time broadening due to motion of the molecules across the laser beam, and some saturation broadening. However, the width of the dip will be considerably narrower than the full Doppler linewidth. For the a_{10} hyperfine component of the R(56) 32-0 band, this width can be less than 1 MHz. A weaker probe beam is sent through the cell in the opposite direction, and is used to lock the laser frequency to that of the resulting narrow transmission peak. The a_{10} hyperfine component is well separated from the other components of the line and is near the center of the overall hyperfine structure, so there is little effect from the wings of the other components.

The results of Ye et al. (2001) for the frequency stability of an iodine-stabilized laser are shown in Figure 2. The iodine vapor pressure was held at 0.8 Pa by cooling a side-tube to -15°C . For laser frequency stabilization work, the use of low sample pressure is important in terms of both minimizing the collision-induced pressure shift and reducing the influence on the baseline by the linear Doppler background. The signal size will of course decrease as the pressure goes down; however, the light path length inside the cell can be extended using a multi-pass strategy or even cavity enhancement. In the current work, the 1.2-m cell length is already

long enough to give adequate signal-to-noise ratio (~ 120 in a 10 kHz bandwidth) without using a build-up cavity. Another important aspect of using lower pressure is that a lower optical power is needed to saturate the transition, because of the reduced collisional broadening. This yields a smaller power-related center frequency shift. The overall reduction of the operational resonance linewidth can to a certain degree compensate for the loss of the signal size since it is the ratio of the linewidth to the signal size that determines the residual rms frequency noise of the stabilized laser. The laser beams were collimated to give a 2 mm mode diameter in the center of the absorption cell, with the probe beam cross-section slightly smaller than the pump beam. With 1.0 mW of saturating power and 0.3 mW for the probe beam, a linewidth of 540 kHz was achieved. The fractional frequency fluctuations (Allan standard deviation) achieved over measurement times τ are given as a function of τ . The results can be described roughly as

$$(\delta\nu/\nu) = 7 \times 10^{-14} / [(\tau)^{0.5}] \quad (11)$$

up to 100 s period, and constant between 100 and 600 s. However, these results were obtained with a 1.2 m long absorption cell and an I_2 vapor pressure of about 15 Pa at 1 °C.

For a space-qualified stable laser, limiting the absorption cell length to about 10 cm appears desirable. This can be done by putting the cell inside a bow-tie cavity, with a path length of about 50 cm and a finesse of roughly 35, as discussed by Ishibashi, Ye and Hall (2002). Also, the I_2 vapor pressure would be reduced to about 0.8 Pa by operating with a side-tube temperature near -15 °C. In this case, the pressure shift coefficient of -3.2 kHz/Pa (Ye et al., 1999) and the temperature-dependent pressure shift slope of about 0.07 Pa/K would give a nominal temperature coefficient of roughly -220 Hz/K. However, some care in the temperature stabilization will be needed, since temperature changes for the rest of the cell can give transient I_2 pressure changes because of residual iodine molecules on the cell walls with different binding energies. But the more compact absorption cell reduces the temperature control problem considerably because of reducing the volume that has to be controlled.

With the type of system discussed above, the calculated frequency instability is about $3.6 \times 10^{-15} / (\tau^{0.5})$ (Ishibashi et al., 2002). A possible alternate approach to using a low-finesse resonant cavity around the cell is to reduce the frequency noise in the NPRO laser below its usual level of about 5 kHz at 1 ms by locking it to a stable resonant cavity for short times. However, until the performance of either type of system has been demonstrated with a space-qualifiable system, it appears desirable to limit our expectations for the performance of iodine-stabilized lasers in space in the near future to something like an Allan standard deviation of 1×10^{-14} or better over almost all of the desired frequency band of 0.0002 to 0.1 Hz, except possibly right at the lower end because of thermal effects at 1 cycle/rev.

Research directed toward the development of a space-qualifiable iodine-stabilized laser system currently is in progress at the Jet Propulsion Laboratory. Some

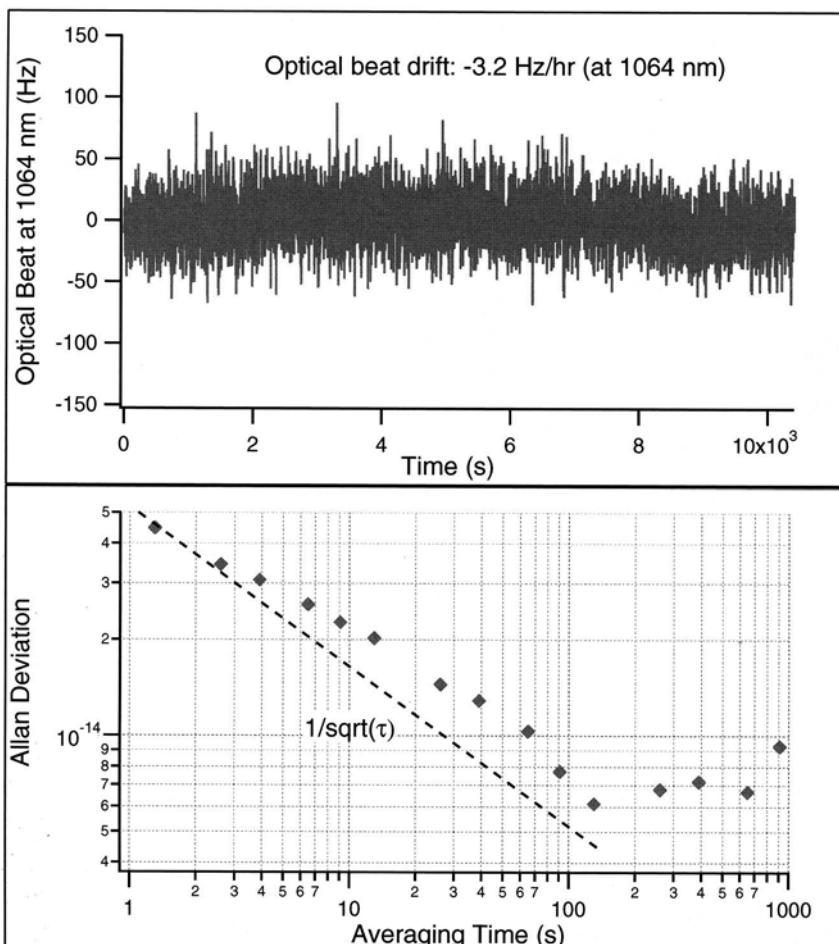


Figure 2. Time record of the beat frequency between the two iodine-stabilized Nd:YAG lasers, with a 1-s frequency-counting gate time. The Allan standard deviation is determined from the beat frequency data.

of the apparatus used in a preliminary laboratory system is shown in Figure 3. The length of the iodine cell shown in the lower part of the picture is 25 cm.

To reach a higher frequency stability, it is useful to explore I_2 transitions with narrower transition linewidths. In fact, I_2 transitions near 514 nm have natural linewidths five times narrower than that at 532 nm, while the signal-to-noise ratios of the recovered sub-Doppler signals are excellent. Recent results (Cheng et al., 2002) indicate that I_2 transitions at 514 nm hold a great promise for future development of high quality, portable optical frequency standards, especially considering the rapid development of all solid state Yb:YAG lasers around 1030 nm.



Figure 3. Preliminary laboratory system for use in research on iodine-stabilized lasers.

3. Conclusions

It appears practical to design and demonstrate iodine-stabilized lasers for use in space measurements of the Earth's gravity field in the near future. Suitable 1064 nm Nd:YAG lasers can be stabilized using only 1 or 2 mW of 2nd harmonic power. The temperature control requirements can be simplified by using quite small iodine absorption cells and fairly low iodine vapor pressure. The expected fractional frequency stability is 1×10^{-14} or better over almost all of the desired frequency range from 0.0002 to 0.1 Hz.

There are substantial advantages to using a stabilized laser of this kind for measuring changes in distances between spacecraft rather than a microwave system. The iodine-stabilized laser system is considerably simpler than any microwave frequency sources of similar stability that have been developed so far. The transmission optics also are simpler for the lasers, and the effects of phase measurement noise are very small. Thus the laser heterodyne measurement systems can be quite compact, and put only modest constraints on the spacecraft design.

Acknowledgements

Part of the research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

- Bender, P.L.: 1992, 'Integrated laser Doppler method for measuring planetary gravity fields,' in From Mars to Greenland: Charting Gravity with Space and Airborne Instruments, IAG Symposium, Vol. 110, Springer-Verlag, pp. 63–72.
- Cheng, W.-Y., Chen, L., Yoon, T.-H., Hall, J. L., and Ye, J.: 2002, 'Sub-Doppler molecular iodine transitions near the dissociation limit (523 to 498 nm),' *Opt. Lett.* **27**, p. 571.
- Colombo, O.L. and Chao, B.F.: 1992, 'Global gravitational change from space in 2001,' IAG Symposium Vol. 112, Potsdam.
- Danzmann, K.V. et al. : 1998, LISA-Laser Interferometer Space Antenna Pre-Phase A Report, 2nd edn., MPQ 233.
- Hammesfahr, A.: 2001 'LISA mission study overview,' *Class. & Quantum Grav.* **18**, pp. 4045-4051.
- Ishibashi, C., Ye, J. and Hall, J.L.: 2002, 'Issues and applications in ultra-sensitive molecular spectroscopy,' in Methods for Ultrasensitive Detection II, C. W. Wilkerson, Jr., Ed., SPIE Vol. 4634 pp. 58–69.
- LISA Study Team: 2000, 'LISA Laser Interferometer Space Antenna: a Cornerstone Mission for the Observation of Gravitational Waves,' ESA-SCI(2000)11, European Space Agency.
- Robertson, D. I., McNamara, P., Ward, H. and Hough, J.: 1997, 'Optics for LISA,' *Class. & Quantum Grav.* **14**, pp. 1575-1577.
- Schumaker, B.L.: 1990, 'Scientific Applications of Frequency-Stabilized Laser Technology in Space,' Jet Propulsion Laboratory Pub. 90-50, Caltech, Pasadena, CA, pp. 133–146.
- Watkins, M.M., Folkner, W.M., Chao, B., and Tapley, B.D.: 2000, 'EX-5: A laser interferometer follow-on to the GRACE mission,' presented at GGG2000, Banff, Canada, 31 July - 5 August, 2000 (unpublished).
- Ye, J., Robertson, L., Picard, S., Ma, L.-S., and Hall, J. L.: 1999, 'Absolute frequency atlas of molecular I_2 lines at 532 nm,' *IEEE Trans. Instrum. Meas.* **48**, pp. 544–549.
- Ye, J., L.-S. Ma, and J. L. Hall: 2001, 'Molecular iodine clock,' *Phys. Rev. Lett.* **87**, pp. 270801/1-4.

Address for Offprints: Scientific Reports Office, JILA, 440 UCB, Boulder, Colorado 80309-0440 USA, sro@jila.colorado.edu

POSSIBLE FUTURE USE OF LASER GRAVITY GRADIOMETERS

P. L. BENDER (pbender@jila.colorado.edu)

JILA, National Institute of Standards and Technology and University of Colorado, Boulder,
Colorado 80309-0440 USA

R. S. NEREM (nerem@colorado.edu)

Colorado Center for Astrodynamics Research and Aerospace Engineering Sciences Department,
University of Colorado, Boulder, Colorado 80309-0431 USA

J. M. WAHR (wahr@longo.colorado.edu)

Cooperative Institute for Research in Environmental Sciences and Physics Department, University
of Colorado, Boulder, Colorado 80309-0390 USA

Received: 11 June 2002; Accepted in final form: 26 September 2002

Abstract. With the GRACE mission under way and the GOCE mission well along in the design process, detailed questions concerning the type of future mission that may follow them have arisen. It is generally agreed that determining the time variations in the Earth's gravity field with as high spatial and temporal resolution as is feasible will be the main driver for such a mission. The possible use of laser heterodyne measurements between separate satellites in such a mission has been discussed by a number of people. The first suggestion of emphasizing time variation measurements in a laser mission was the TIDES concept presented in 1992 by Colombo and Chao. Then, in 2000, a GRACE Follow-On mission using laser measurements between two drag-free satellites was discussed by Watkins et al. (2000).

More recently, the possibility of utilizing laser measurements between more than two satellites in order to determine two or more components of the gravity gradient tensor simultaneously has been proposed by Balmino. This approach may be desirable in order to reduce the aliasing of time variations between geopotential terms of different degree and order, as well as to improve the resolution in longitude, despite the cost of the additional satellites. In this paper, we discuss specific possible mission geometries for measuring the two diagonal in-plane components of the gravity gradient tensor simultaneously. This could be done, for example, by laser heterodyne measurements between two pairs of satellites in coplanar and nearly polar orbits.

1. Introduction

The idea of using laser measurements between satellites for mapping the Earth's gravity field was a natural development from studies of similar missions using microwave measurements (MBB, 1978; Bender, 1985; Schumaker, 1990; Bender, 1992). The first proposal for detailed determinations of time variations in the Earth's gravity field by laser heterodyne measurements was made by Colombo and Chao (1992). In their proposed TIDES mission, the spacecraft were to be as simple as possible, and thus they were not drag-free.



Space Science Reviews **108:** 385–392, 2003.

© 2003 Kluwer Academic Publishers.

The GRACE mission was proposed in 1996 in response to a competitive NASA opportunity for which it did not seem feasible to include either drag-free systems or laser heterodyne technology. The measurement of time-variation in the gravity field over a five year period is a major objective. Then ESA chose the GOCE gravity gradiometer mission in 1999, with emphasis on determining the short wavelength part of the gravity field more accurately by measuring three components of the gravity gradient tensor simultaneously at low altitude in a drag-free satellite.

For the future, it seems desirable to consider what kinds of laser heterodyne measurements between drag-free satellites could be candidates for a follow-on to the GRACE and GOCE missions in order to continue and improve measurements of time variations in the Earth's gravity field. Quite high spatial resolution also is expected to be needed, since for example gravity changes due to hydrological variations can occur quite rapidly over the shortest distance scales observable from space. We will consider here only measurement uncertainties, rather than the more difficult issue of how well time variations in gravity due to different causes such as atmospheric, oceanographic and hydrological changes can be unraveled from each other. Also, only errors due to laser frequency noise, accelerometer noise, and noise in possible control forces applied to the test masses in the accelerometers will be considered. Shot noise and beam pointing noise in measuring the distances between the satellites are expected to be negligible.

2. One-Component Laser Gravity Gradiometer

One possible future follow-on to the GRACE mission has been discussed by Watkins et al. (2000). It would include drag-free operation of both satellites, as well as a laser link between them. We assume as an example nearly polar orbits, with perhaps 310 km altitude and 100 km spacecraft separation. Separate ephemeris terms can be calculated for the midpoint of the baseline between the two satellites and for the baseline vector between them. The information obtained from changes in the length of the baseline vector would be much like that obtained from a gradiometer measuring the component Γ_{yy} of the local gravity gradient tensor, where y is the along-track component of the baseline vector. Thus we will refer to such a mission as a one-component laser gravity gradiometer mission. This choice of terminology is made because we believe it makes communicating the type and accuracy of the gravity field information that can be obtained to Earth scientists easier.

The 310 km altitude suggested is about 60 km higher than is planned for the GOCE three-component gravity gradiometer mission. The reasons for this are mainly the roughly 10 year lifetime desired for a future laser mission and the need to keep the propulsion part of the drag-free system as simple as possible in order to minimize the associated power and fuel requirements.

An Allan standard deviation of 1×10^{-14} independent of averaging time is assumed for the laser (Bender et al., 2002). This corresponds to a spectral amplitude of the fractional laser frequency variations of

$$(\delta\nu)/(\nu) = (0.85 \times 10^{-14})(1\text{Hz}/f)^{0.5}/\sqrt{\text{Hz}}, \quad (1)$$

where f is the frequency of the variations. The corresponding fractional variations in the laser wavelength, and thus in the measured baseline length, are the same. With the assumed 100 km spacecraft separation, this gives variations in the apparent baseline length of

$$\delta L = (0.85 \times 10^{-9})(1\text{Hz}/f)^{0.5}\text{m}/\sqrt{\text{Hz}}. \quad (2)$$

It appears feasible to make all other contributions to measuring changes in the baseline length less than this for f between 0.1 and 0.001 Hz.

Multiplying δL by $(2\pi f)^2$ gives the variations in the apparent relative along-track acceleration due to the laser frequency variations of

$$\delta a_L = [(2\pi)^2](0.85 \times 10^{-9})(f/1\text{Hz})^{1.5}\text{m/s}^2/\sqrt{\text{Hz}}. \quad (3)$$

This corresponds to a gravity gradient noise level of $[1.1 \times 10^{-5}](f/0.1\text{Hz})^{1.5}\text{E}/\sqrt{\text{Hz}}$. For an independent accelerometer noise level of

$$\delta a_A = 1 \times 10^{-12}\text{m/s}^2/\sqrt{\text{Hz}} \quad (4)$$

for each of the two accelerometers, δa_L would be $2^{0.5} \times \delta a_A$ at $f = 1.2$ mHz. Thus the laser noise would dominate strongly at higher frequencies. This emphasizes the importance of minimizing the laser frequency noise in order to achieve as high spatial resolution as possible.

A potential limitation of one-component gravity gradiometer data for determining short wavelength features in the Earth's gravity field and their time variation is the difficulty of separating time variations from spatial variations (Thompson et al., 2000). If, e.g., 30 day repeat orbits are used, it may be a few days between times when equatorial crossings come within roughly 1° in longitude of each other. Thus, if the data is simply averaged over 20 or 30 day blocks, time variations over periods of a few days can be interpreted as short wavelength spatial variations in the East-West direction.

As an extreme example, assume no accelerations are observed for one South-to-North pass, but there are time variations of low and medium degree potential terms that are not nearly sectorials before measurements are made over another ground track about 1° away. The results could be interpreted as indicating short wavelength spatial variations in the potential in the East-West direction if care is not taken in the analysis of the data.

Clearly the effect of aliasing of temporal and spatial variations will tend to average out over longer times. However, aliasing may be a substantial limitation if good resolution is desired in both time and space. An approach that is being

tried to reduce this problem for the GRACE mission is to solve for a low degree gravity field every day or every few days, and then to remove the time-varying part before solving for the higher degree parts of the field (Thompson et al., 2000). Also, there is of course information about the East-West potential variations contained in the differences between the data for ground tracks for South-North and North-South passes, as shown in early studies of Breakwell (1979), even though the angle between the ground tracks near the equator is only about $1/8\text{ rad}$ for polar orbits. Requiring consistency of the potential at such track crossings, as well as in the polar regions, may suggest ways of reducing the effect of time variations. Experience with the GRACE data is expected to be valuable in evaluating the effectiveness of many different approaches to reducing aliasing.

3. Possible Two-Component Laser Gravity Gradiometer

In view of the difficulty of determining the East-West variations of the gravity field with the highest accuracy with Γ_{yy} only, even if there were no time variations, it has been suspected that there would be substantial advantages to measuring two or more gravity gradient components simultaneously. Possible satellite configurations for making such measurements with satellite-to-satellite distance measurements have been discussed by Balmino (2002). The second component can be either Γ_{xx} or Γ_{zz} , since the trace of Γ is zero. Also, two different linear combinations of Γ_{yy} and one of these other diagonal components can be measured.

One attractive possibility for a two-component laser gravity gradiometer uses an array of four satellites in what has been called a cartwheel configuration (see Figure 1). The orbits are coplanar and nearly polar, with equal semi-major axes, eccentricities, and mean longitudes. The arguments of perigee are

$$\omega_i = \omega_1 - (i - 1) \times 90^\circ \quad (5)$$

for $i = 1, \dots, 4$. In a reference frame that rotates once per revolution around the normal to the orbit plane, the satellites follow oval orbits around the center of the array.

As an example, we take $e = 0.00375$, so that the baselines between nonadjacent satellites will vary in length from about 50 to 100 km at twice the orbital frequency. Laser beams would be sent both ways between satellites S-1 and S-3 along the baseline L1, and similarly for L2. The Doppler shifts for the signals received at each satellite would be up to 60 MHz. Also, the angles between each baseline and a vector rotating at once per rev will vary by $\pm 20^\circ$ at a rate of once per rev. The Doppler shifts and the non-uniformly rotating baselines are the main challenges for the cartwheel configuration.

We assume for our example that the maximum thrust required for drag compensation is very roughly $1 \times 10^{-3} N$, and that maximum power required is 150 W per satellite. This means that there is a substantial advantage to Sun-synchronous

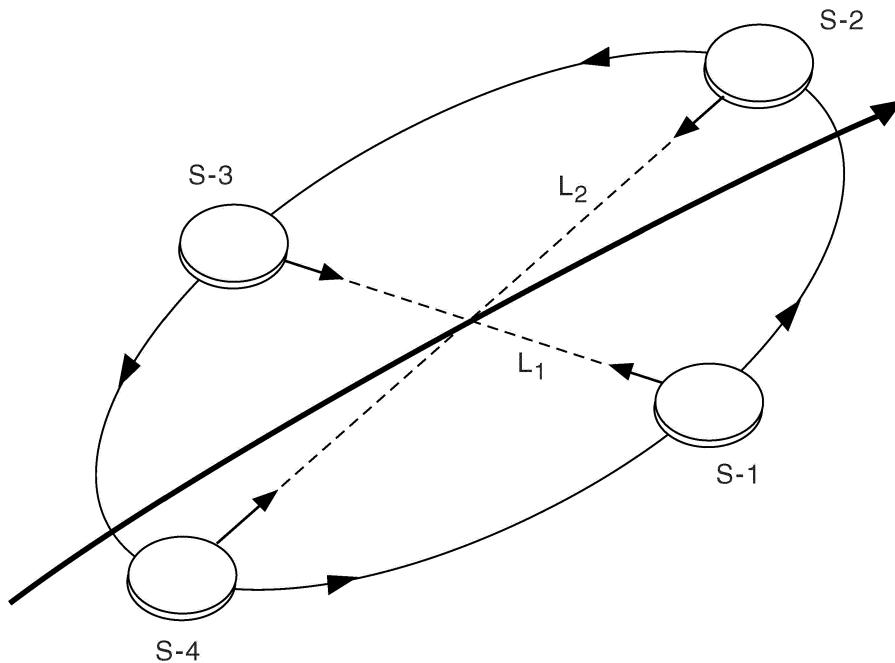


Figure 1. The heavy curve indicates the motion of the center of the satellite array around the Earth in a nearly polar plane. The light curve shows the relative motion of the satellites around an oval path once per orbital revolution. Laser beams are sent both ways between opposite pairs of satellites.

operation, to minimize the area of solar panels needed and to simplify the geometry. For this approach, the inclination is about 96.8°, as for GOCE, and the main disadvantages are the aliasing of some tidal frequencies and the lack of coverage near the poles. We estimate that a cylindrical satellite about 1.3 m in diameter and much less in height, with a mass of about 100 kg, might be suitable. If the axis of the cylinder is perpendicular to the orbit plane and a body-fixed perpendicular axis is kept pointed along the baseline vector L₁ or L₂, the required maximum torque on the satellite is roughly 2×10^{-5} Nm.

With the above approach, if a rectangular test mass is used as in GRACE and GOCE, the maximum torque required to keep the test mass oriented with a face perpendicular to a baseline vector is very roughly 10^{-10} Nm. If this torque is applied by a pair of capacitive forceps 1 cm off axis, each forceps must provide a

force of up to 5×10^{-9} N. As a guess, the stability of the resulting translational force might be $2 \times 10^{-5}(1\text{Hz}/f)/\sqrt{\text{Hz}}$ as large at frequencies of 0.2 to 100 mHz, and mainly transverse to the most sensitive axis. The resulting acceleration noise of roughly $10^{-13}(1\text{Hz}/f)\text{m/s}^2/\sqrt{\text{Hz}}$ would be of concern if it were along the most sensitive axis, since it would be the dominant noise source below about 6 mHz. The design of the accelerometer in such a way that a fairly large torque could be applied to the test mass without substantially compromising the accelerometer performance at low frequencies thus will be an important goal if this approach is used.

An alternative approach is to use a spinning proof mass, as suggested by De Bra (1998) and Lange (2001, 2002) for other applications. The test mass would be spun around a convenient axis at a frequency considerably above the measurement frequency band. The effects of departures of the surface from a sphere and the offset of the center of figure from the center of mass can be avoided by proper design of the system.

Still another approach would be to provide a small optical assembly that could be rotated with respect to the satellite to keep it pointed along the desired L1 or L2 baseline, but to let the spacecraft rotate uniformly at a rate of one cycle per orbital revolution. This could be done with either a rectangular test mass or a spherical one. With this approach, the number of satellites could be reduced to three, with two separately pointed optical assemblies per satellite. This would give some redundancy, since up to two of the six optical assemblies could be lost but measurements along two sides of the triangle could still be made, as long as the two lost optical assemblies were not on the same satellite. This is similar to what is planned for the LISA mission (LISA Study Team, 2000), but there the required angle change is only about 0.6° in half a year or so, rather than roughly 20° in half a revolution.

Finally, if the tidal aliasing and the loss of gravity field information at the poles were not acceptable, a more nearly polar orbit could be chosen. The main result would be less efficient use of the solar illumination to generate power on the satellite, and thus the need for a larger solar panel area.

An important point that has not yet been investigated carefully is the degree of difficulty of making the necessary Doppler phase or frequency measurements when the Doppler shift is as large as 60 MHz in amplitude and changing at a rate of twice per rev. It probably would be necessary to use a number-controlled oscillator (frequency synthesizer) to generate a reference signal that the laser heterodyne signal could be beat against to get it down to a more convenient frequency to work with.

One approach is to only change the reference frequency occasionally, and then to sample the reduced frequency with a rapid sampler at a 20 MHz or 40 MHz rate, as is done in high performance satellite GPS receivers. The rest of the analysis can then be done in software. Another approach is to put much more of the burden on the frequency synthesizer and keep the reduced output frequency down to perhaps

a few kHz. In this case, conventional timing of the resulting zero-crossings can be used to determine the changes in satellite separations. The rapid sampling approach was suggested for the LISA mission by Folkner (1995), and the zero-crossing approach for LISA is being investigated by Pollack and Stebbins (see Jennrich et al., 2001).

At first glance, it may seem desirable to also consider nearly polar orbits similar in principle to those for the LISA mission (Folkner et al., 1997), but geocentric instead of heliocentric. These orbits have small inclination with respect to a reference plane, but the longitudes of the nodes are quite different, so the orbits are not coplanar. The eccentricities also are small, but the difference between the longitudes of perigee and the node would have to be close to 90° for each satellite. Since the precession of the node is small for nearly polar orbits but the precession of perigee is not, the conditions needed for nearly polar LISA-like geocentric orbits cannot be maintained without strong forces being applied to the test masses in the spacecraft. A serious concern is that applying strong forces will lead to noise forces in the measurement frequency band also acting on the test masses. This is the reason for choosing the cartwheel configuration for the example described in this section.

4. Conclusions

Substantial further improvements in the determination of time variations in the Earth's gravity field in space seem feasible. For this purpose, it appears useful to consider laser gravity gradiometers that measure either one or two components of the gravity gradient tensor. The number of satellites required is between two and four. However, it appears that the individual spacecraft can be made relatively simple, with the main size determining factor being the area of solar panels needed for propulsion to overcome the drag at the desired fairly low measurement altitude. Further technology development on both highly stable lasers and suitable ion propulsion systems are certainly desirable.

With cartwheel type orbits for two-component laser gravity gradiometers, Sun-synchronous orbits like that for GOCE would be simpler, but polar orbits also appear to be feasible. The use of rectangular test masses in the accelerometers requires applying substantial torques to keep the orientation correct, but seems quite possible. However, the use of spherical proof masses also deserves consideration. The feasibility of making accurate phase measurements on laser heterodyne signals with rapidly changing Doppler shifts needs to be checked. And finally, detailed studies of the results achievable with one-component and two-component laser gravity gradiometers are needed, with particular attention to the problem of aliasing low degree time variations into higher degree gravity field terms.

Acknowledgements

It is a pleasure to thank Georges Balmino, Gerhard Beutler, Reiner Rummel and Mike Watkins for useful information related to this paper. This work was supported in part by NASA grants NAG5-10259, NAG5-5171, and NAG5-7703 to the University of Colorado.

References

- Balmino, G.: 2002, private communication.
- Bender, P. L.: 1985, 'Laser Doppler experiment option for a 200-250 km altitude,' Geopotential Research Mission (abstract), EOS, Trans. Amer. Geophys. Union **66**, 242.
- Bender, P. L.: 1992, 'Integrated laser Doppler method for measuring planetary gravity fields,' in From Mars to Greenland: Charting Gravity With Space and Airborne Instruments, IAG Symposium No. 110, Springer-Verlag, 63-72.
- Bender, P. L., Hall, J. L., Ye, J., and Klipstein, W. M.: 2002, 'Satellite- satellite laser links for future gravity missions,' in *Space Sci. Rev.*, this volume.
- Breakwell, J. V.: 1979, 'Satellite determination of short wavelength gravity variations,' *J. Astronaut. Sci.* **27**, 329-344.
- Colombo, O. L. and Chao, B. F.: 1992, 'Global gravitational change from space in 2001,' IAG Symp. 112, Potsdam.
- De Bra, D. B.: 1998, 'Design considerations for drag free satellites,' in Laser Interferometer Space Antenna, Amer. Inst. Phys. Conf. Proc. 456, W. M. Folkner, Ed., AIP, Woodbury, N.Y., 199-206.
- Folkner, W. M.: 1995, private communication.
- Folkner, W. M., Hechler, F., Sweetser, T. H., Vincent, M. A., and Bender, P. L.: 1997, 'LISA orbit selection and stability,' *Class. & Quantum Grav.* **14**, 1405-1410.
- Jennrich, O., Stebbins, R. T., Bender, P. L., and Pollack, S.: 2001, 'Demonstration of the LISA phase measurement principle,' *Class. & Quantum Grav.* **18**, 4159-4164.
- Lange, B.: 2001, 'Managing spherical proof masses in drag-free satellites with application to the LISA experiment,' *Class. & Quantum Grav.* **18**, 4153-4158.
- Lange, B.: 2002, 'Drag-free performance in a LISA mission with spherical proof masses,' *Class. & Quantum Grav.* **19**, 1739-1743.
- LISA Study Team: 2000, 'LISA Laser Interferometer Space Antenna: a Cornerstone Mission for the Observation of Gravitational Waves,' ESA-SCI(2000)11, European Space Agency.
- MBB: 1978, 'SLALOM Mission/System Definition - Final Report,' European Space Agency Contract No. 3483/78/F/DK(SC), Messerschmitt-Boelkow-Blohm GMBH, Ottobrunn/Muenchen.
- Schumaker, B. L.: 1990, 'Scientific Applications of Frequency-Stabilized Laser Technology in Space,' Jet Propulsion Laboratory Pub. 90-50, Caltech, Pasadena, CA, pp. 133-146.
- Thompson, P. F., Bettadpur, S. V., Kim, J., and Watkins, M. M.: 2000, 'Short period time variations in the gravity field and their impact on GRACE science,' Trans. of the AGU (EOS) **81**, F310.
- Watkins, M. M., Folkner, W. M., Chao, B., and Tapley, B. D.: 2000, 'EX-5: A laser interferometer follow-on to the GRACE mission,' presented at GGG2000, Banff, Canada, 31 July - 5 August, 2000 (unpublished).

MICROSCOPE INSTRUMENT DEVELOPMENT, LESSONS FOR GOCE

PIERRE TOUBOUL

*Physical Instrumentation Department, ONERA, F-92322 Chatillon Cedex, France
(touboul@onera.fr)*

Received: 4 November 2002; Accepted in final form: 27 November 2002

Abstract. Two space missions are presently under development with payload based on ultra-sensitive electrostatic accelerometers. The GOCE mission takes advantage of a three axis gradiometer accommodated in a very stable thermal case on board a drag-free satellite orbiting at a very low altitude of 250 km. This ESA mission will perform the very highly accurate mapping of the Earth gravity field with a geographical resolution of 100 km. The MICROSCOPE mission is devoted to the test of the “Universality of free fall” in view of the verification of the Einstein Equivalence Principle (EP) and of the search of a new interaction. The MICROSCOPE instrument is composed of two pairs of differential electrostatic accelerometers and the accelerometer proof-masses are the bodies of the EP test. The satellite is also a drag-free satellite exhibiting a fine attitude control and in a certain way, each differential accelerometer is a one axis gradiometer with an arm of quite null length. The development of this instrument much interests the definition and the evaluation of the sensor cores of the gradiometer. The in flight calibration process of both instruments is also very similar. Lessons from these parallel developments are presented.

1. Introduction

In the last decade, several space missions have involved very high sensitive accelerometers with a resolution better than 1 nano- g (1 g = Earth mean gravity field on ground). In the SpaceLab flying on board the COLUMBIA shuttle, the survey at low frequency (down to 0.1 mHz) of the residual acceleration level has been performed twice for dedicated micro-gravity experiment missions with electrostatic three-axes accelerometers (Nati *et al.*, 1994; McPherson *et al.*, 1999) down to a level of one nano- g .

The CHAMP and the GRACE satellites have been launched respectively in July 2000 and March 2002 and carry at their centre of mass such an instrument for the fine measurement of the surface forces acting on the satellites (Reigber *et al.*, 1996; Reigber *et al.*, 2002; JPL, GRACE, 1998; Tapley and Reigber, 2002). CHAMP is the precursor of this new generation of space geodesy missions, all considering on board measurements of the acceleration or gravity field. The two GRACE satellites are flying on the same orbit, injected at 480 km with a 98° inclination and a separation distance of about 230 km. The relative velocity of the satellite is as low as 1 km per day. The fine spacecraft trajectography is performed by the on board GPS receivers and the relative motion of the satellite is measured by one accurate micro-wave tracking, leading, with the accelerometer outputs together



Space Science Reviews **108:** 393–408, 2003.

© 2003 Kluwer Academic Publishers.

to the recovery of the monthly variations of the Earth gravity field. The required performance for the accelerometer is 10^{-10} ms^{-2} over 1 second integration period.

Two new space missions are now under development respectively in ESA and CNES, requiring sensors with improved performance by a factor one hundred. The GOCE mission takes advantage of a three axis gradiometer accommodated on board a drag-free satellite orbiting at a very low altitude of 250 km; the atmospheric drag of the satellite is so compensated by the continuous and controlled thrust of the electrical propulsion system (ESA, 1999). This ESA mission will perform the very accurate mapping of the Earth gravity field with a geographical resolution of 100 km. Each of the three one axis gradiometer constituting the instrument, is composed of a very steady structure and two highly sensitive inertial sensors mounted at the two ends, half a meter distance from each other.

The MICROSCOPE mission, also scheduled for a launch in 2006, is devoted to the test of the “Universality of free fall” in view of the verification of the Einstein Equivalence Principle (EP) or of the search of a new interaction (Touboul *et al.*, 2002). The MICROSCOPE instrument is composed of two pairs of differential electrostatic accelerometers and the accelerometer proof-masses are the bodies of the EP test. The satellite is also a drag-free satellite exhibiting a fine attitude control and in a certain way, each differential accelerometer is a one axis gradiometer with an arm of quite null length. The expression of the performed measurements are very similar in both cases but in MICROSCOPE the gravity gradient is the disturbing signal while in GOCE, it is the signal to be measured. The developments of both instruments are also comparable because of the same concept of operation and of the same approach to increase the resolution from the already existing accelerometers presently in orbit. This is in particular possible because of the limited range of operation thanks to the vibration soft environment provided by the drag-free satellite and of the very stable thermal environment inside the controlled instrument case. Some lessons form the MICROSCOPE instrument development are thus pointed out to assess the selected configuration for the GOCE gradiometer.

2. MICROSCOPE Mission

More accurate experiments confirming the equivalence between inertial mass and gravitational mass represent important verifications of the relativity theory of gravitation and other metric theories, which postulate this principle. The search of a new interaction (with extra mass-less scalar field in particular) expected by the string theories in view of a quantum gravity theory is also a great and recent motivation for ground and space experiments in addition to the determination with better accuracy of the Post Newtonian coefficients (Damour *et al.*, 2002; Lämmerzahl *et al.*, 2000).

The MICROSCOPE mission (MICROSatellite pour l’Observation du Principe d’Equivalence), has been selected by CNES with three other micro-satellite mis-

sions devoted to Earth observation DEMETER, PARASOL and PICARD (CNES, 2002). This fundamental physics space experiment aims at the EP test with an expected accuracy of at least 10^{-15} , more than two orders of magnitude better than the last ground tests performed with torsion pendulum and limited by the Earth gravity gradient, the seismic noise and the disturbances induced by human activities (Fischbach and Talmadge, 1998).

Recent altimetry and geodesy missions lead to the global and accurate determination of the Earth gravity field expressed with spherical harmonics series and to the production of dedicated mathematical and computational tools for the simulation of accurate orbital motion. Micro-satellites with a mass limited to 150 kg maximum are now available with low cost launch opportunities like ARIANE V secondary small passengers or Russian or Ukrainian launchers. In addition, the experience of the measurement of weak accelerations by electrostatic accelerometers performed on board the CHAMP and the GRACE satellites confirm the way of a fast and cheap space experiment devoted to the EP test and preparing more ambitious other missions in fundamental physics (Touboul *et al.*, 1998; Touboul, 2000).

The MICROSCOPE experiment consists in maintaining two test masses made of different materials along the same orbit. The relative motion of the two masses is finely measured and electrostatic fields are generated all around each mass to apply well controlled electrostatic pressures. The symmetry of the configuration is of peculiar importance. The two masses are cylindrical and concentric (their centre of mass at the same point) in such a way that they are submitted to the same Earth gravity field. In case of no violation of the EP, the electrostatic fields, controlled around the mass generate the same acceleration. These electrostatic accelerations are finely measured along the Earth pointing axis that is projected along the three instrument axes according to the satellite attitude. So, when the satellite is inertial pointing or orbiting about the normal axis to the orbital plane, the EP signal to be detected will be modulated respectively at the orbital frequency or at the sum (or difference according to the sign of the satellite rotation) of the spin and orbital frequency, *i. e.* between 1.7×10^{-4} Hz and a few 10^{-3} Hz.

3. MICROSCOPE Instrument and Accommodation

The satellite payload is composed in fact of two quite identical differential electrostatic accelerometers, each of the two pairs of masses being the two proof-masses of two accelerometers. The masses are made of the same Platinum Rhodium alloy for the first one, dedicated to assess the accuracy of the experiment: for a perfect experiment, the outputs of the two accelerometers are identical. The materials are Platinum Rhodium alloy and Titanium or Copper alloy for the second one. Then, in view of suppressing the systematic errors, the experiment logic relies on the double

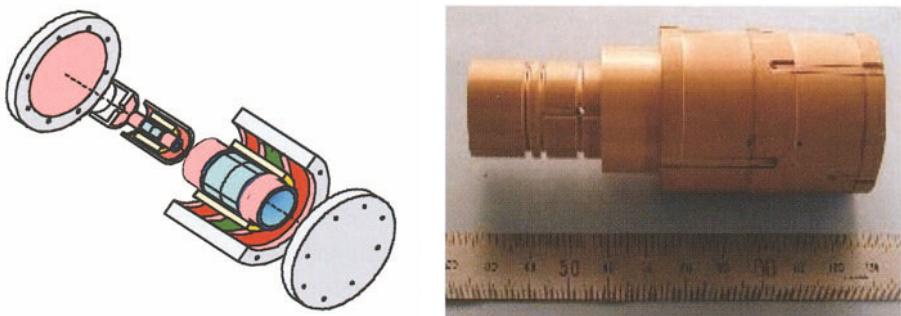


Figure 1. Differential accelerometer configuration: each of the two masses, in yellow, are integrated inside a silica core composed of the inner and the outer cylinders which carry the electrodes: radial in blue on the internal cylinder and axial in red on the external one (left picture). Prototype of the inner inertial sensor (right picture).

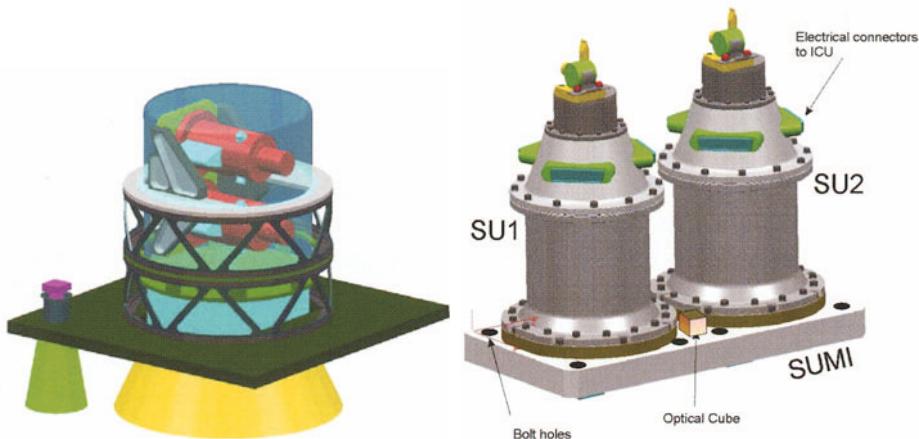


Figure 2. Accommodation of the two tight housings (right) inside the double-stage thermal insulated case of the payload (left). The yellow cone is turned normal to the orbital plane, to the opposite of the Sun. It protects the external radiator of the payload case from the Earth radiation fluctuations in order to increase its temperature stability and so the temperature of the payload especially at the f_{EP} frequency.

comparison of the outputs of two pairs of electrostatic accelerometers. The weight of the masses ranges from 0.4 kg up to 1.7 kg.

One differential accelerometer is composed of two concentric electrostatic inertial sensors (see Figure 1) with fused silica cylindrical cage carrying sets of electrodes engraved in the cage gold coatings all around the mass. The configuration of the electrode set presents an axial and a radial symmetry.

Pairs of electrodes are used for the capacitive sensing of the mass position and attitude along the three axes (Josselin *et al.*, 1999). The same electrodes are used to generate the electrostatic fields around the mass for the control of its motion. The eight quadrant electrodes concern the radial translations and rotations. The

two cylindrical sensing electrodes located at the ends of the test-mass are used for the axial direction. The rotation of the mass about the axial direction is measured through dedicated flat areas on the mass and external electrodes in regard.

The cylindrical configuration allows the centering of the two masses at the same point to suppress the effect of the satellite or the Earth gravity gradients. It is also optimised for the reduction of the electrostatic stiffness and damping associated to the mass motion along the axial direction to the benefit of performance of the one major axis along which the EP test is performed. The mean electrostatic force applied on both masses is in fact nullified by the satellite drag compensation system that acts through a servo-loop on the thrusters of the satellite which so follows the masses.

The position and the attitude of the masses can be finely adjusted by offsetting the six servo-loops which performs from the capacitive sensing the electrostatic control of the mass with respect to the instrument frame. The instrument sensitivity to the mass relative position will be verified during the calibration phase as well as the rejection rate of the Earth gravity gradient signal depending on it.

As shown in Figure 2 (right), the instrument cores are integrated in tight vacuum housings that provide also thermal insulation and magnetic shielding. The two housings, mounted on a rigid structure, are then accommodated inside a double stage thermal case. The satellite external structure, in green on the left picture, carries the star sensors that have to be steadily aligned with the inertial sensor (better than 1 arc min): the control of the satellite attitude is performed from the complementary pointing measures provided by the star sensors and the angular acceleration measures provided by the inertial sensors. The passive thermal control of the sensor units and the associated electronics must meet the requirements expressed in Table I.

These requirements are deduced from the accelerometer thermal sensitivity considering the thermal expansion of the mechanical parts, the sensitivities of the electronics components, the radiation pressures and the radiometric force applied on the masses, plus other minor effects. Obviously, these effects have to be considered about the EP signal frequency, f_{EP} , for the MICROSCOPE mission, and then expressed in power spectral density or at the EP signal frequency (tone). Not only the sine but also the phase of the eventual violation signal is well known because it has to correspond to the Earth gravity monopole direction. Thus, random perturbations can be filtered out through heterodyne detection. This will not be the case of the GOCE mission for which the gravity signal variations to be measured extends in a frequency bandwidth up to 0.1 Hz.

The differential accelerometer housings are mounted near the satellite centre of mass but contrarily to missions like CHAMP or GRACE, no stringent requirement with millimeter accuracy is demanded. In these previous missions, the accelerometer outputs are representative to the external surface forces acting on the satellite and exploited to analyse the orbital motion of the satellite centre of mass: centering the instrument is required to reject the impact of the centrifugal and the angular

TABLE I
 Comparison of the thermal environment requirements for the instrument sensor core and the electronics units in case of the MICROSCOPE and the GOCE missions. Power Spectral Density (PSD) and tone variations are defined at the interface of the mounting plane. Each unit benefits in addition from its own thermal insulation and inertia.

	MICROSCOPE Electronics unit	MICROSCOPE Mechanics unit	GOCE Electronics unit	GOCE Mechanics unit
Operating temperature	+10°C to +50°C	+20°C to +40°C	+20°C to +30°C	+20°C to +25°C
Thermal variations:				
PSD	$2 \text{ K Hz}^{-1/2}$ (about $f_{EP} \sim 10^{-3} \text{ Hz}$)	$200 \text{ mK Hz}^{-1/2}$ (about $f_{EP} \sim 10^{-3} \text{ Hz}$)	$10 \text{ mK Hz}^{-1/2}$ ($5 \times 10^{-3} \text{ Hz}$ -0.1 Hz)	$5 \text{ mK Hz}^{-1/2}$ ($5 \times 10^{-3} \text{ Hz}$ -0.1 Hz)
Tone (sine at f_{EP})	6 mK	0.6 mK	Not applicable	Not applicable
Thermal Gradients:				
PSD	Not applicable	$2 \text{ K/m Hz}^{-1/2}$ (about $f_{EP} \sim 10^{-3} \text{ Hz}$)	Not applicable	$125 \text{ mK/m Hz}^{-1/2}$
Tone (sine at f_{EP})	Not applicable	6 mK/m	Not applicable	Not applicable

accelerations of the satellite and of the gravity gradient. In the case of the MICROSCOPE mission, as well as of the GOCE one, the drag free point of the satellite (*i.e.* falling around the Earth) is defined by the satellite torque and force compensation system. And this point is defined from the accelerometer outputs themselves, so independently of the satellite centre of mass but of the accelerometer proof-masses.

4. MICROSCOPE Differential Measurement

The measure provided by one differential accelerometer can be expressed from the equation of the two mass motions:

$$m_{I_A} (\ddot{X}_A + \ddot{x}_A) - m_{g_A} g_A = F_A + F_{P_A}$$

with I , g and A (or B later), the inertial or the gravitational mass A (or B), X the motion of the instrument frame (or of the satellite) with respect to the inertial frame, x the motion of the proof-mass with respect to the instrument frame,

$$g_i = -\frac{\mu}{r^2} \frac{(X_i + x_i)}{r}$$

gravity field (2) and

$$T_{ij} = -\frac{\mu}{r^3} \left(\delta_{ij} - \frac{3(X_i + x_i)}{r} \frac{(X_j + x_j)}{r} \right)$$

gravity gradient tensor (3) with $\mu \approx 4 \times 10^{14} \text{ m}^3 \text{s}^{-2}$, the Earth gravitational constant and r the distance to the geocentre, $r = \left(\sum_{i=1}^3 (X_i + x_i)^2 \right)^{1/2}$ ($X_i + x_i$) the rectangular co-ordinates of the current point in Earth's fixed orthonormal frame; δ_{ij} is the Kronecker symbol; g_A the Earth gravitational field integrated over the mass volume A , F_A and F_{P_A} , are the forces respectively applied by the electrostatic suspension or by the disturbing sources.

The electrical signal delivered by each inertial sensor is representative of the electrostatic force F_A :

$$\hat{F}_A = (I + K_A) F_A + E(F_A) + E_{n_A}$$

with I and K_A , respectively the identity and the sensitivity matrix, $E(F_A)$ and E_{n_A} , respectively the nonlinearity and the noise of the measurement system.

Then, by performing the difference of the inertial sensor outputs:

$$\begin{aligned}
 \frac{\hat{F}_A}{m_{I_A}} - \frac{\hat{F}_B}{m_{I_B}} &\approx + (K_A - K_B) \frac{\ddot{X}_A + \ddot{x}_A + \ddot{X}_B + \ddot{x}_B}{2} + \left(I + \frac{K_A + K_B}{2} \right) \\
 &\quad \times (\ddot{X}_A - \ddot{X}_B) + \left(I + \frac{K_A + K_B}{2} \right) \\
 &\quad \times \{ (\ddot{x}_A - \ddot{x}_B) + 2\Omega (\dot{x}_A - \dot{x}_B) + (\Omega\Omega + \dot{\Omega})(x_A - x_B) \} \\
 &- \frac{\hat{F}_{p_A}}{m_{I_A}} + \frac{\hat{F}_{p_B}}{m_{I_B}} + \frac{E(F_A)}{m_{I_A}} - \frac{E(F_B)}{m_{I_B}} + \frac{E_{n_A}}{m_{I_A}} - \frac{E_{n_B}}{m_{I_B}} \\
 &- \frac{1}{2} \left(\frac{m_{g_A}}{m_{I_A}} + \frac{m_{g_B}}{m_{I_B}} \right) (g_A - g_B) \\
 &- \left(\frac{m_{g_A}}{m_{I_A}} - \frac{m_{g_B}}{m_{I_B}} \right) \left(\frac{g_A + g_B}{2} \right)
 \end{aligned} \tag{1}$$

The first term of the first line corresponds to the common mode acceleration not fully suppressed in the difference because of the non equal sensitivity matrices. The second and third terms corresponds respectively to the relative residual motion of the masses (limited by the stability of the instrument assembly) and to the effect of the instrument attitude variations, Ω being the angular velocity about the drag free point (time derivatives noted \circ are performed in the rotating frame).

The fourth line depends only on the inertial sensor noises and defects. Motions of the test-masses are measured along the axis of revolution (peculiar sensitive axis) with a resolution of 6×10^{-10} m/ $\sqrt{\text{Hz}}$ and controlled with a resolution around 1.5×10^{-12} m/s $^2/\sqrt{\text{Hz}}$ for an operating range reduced to 5×10^{-7} m/s 2 . The first source of limitation of the resolution at the frequency f_{EP} is the fluctuation-dissipation noise induced by the proof-mass motion damping due to the thin gold wire used for the charge control of the mass:

$$\Gamma_{\text{wire}}^2 = \left(\frac{1}{m} \sqrt{4k_b T \frac{k_{\text{wire}}}{2\pi f_{EP} Q_{\text{wire}}}} \right)^2 (\text{ms}^{-2})^2 / \text{Hz}$$

with Q_{wire} measured greater than 100 and k_{wire} lower than 5×10^{-6} N/m (Willemont and Touboul, 1999). The same charge control is performed for the GOCE accelerometer mass, where the same limitation at lower frequency is encountered. The fourth term corresponds to the gravity gradient signal and the last one to the eventual EP signal.

The atmospheric and thermal drag and torque of the satellite as well as others are actively compensated by the continuous controlled variations of the thrust of the electrical propulsion such that the satellite follows the two test masses in their gravitational motion with residual motion fluctuations of 3×10^{-10} m/s $^2/\sqrt{\text{Hz}}$ in translation and 10^{-8} rad/s $^2/\sqrt{\text{Hz}}$ in rotation. The propulsion system allows also a fine calibration of the instrument by generating well known kinematic accelerations

in all six degrees of freedom. This allows to verify the sensitive axis alignments and the matching of the instrument sensitivities necessary to reject the common kinematic acceleration fields ($K_{A_{ij}} - K_{B_{ij}} < 3 \times 10^{-4}$). The same approach is considered in the GOCE mission except that the drag-free point is at the gradiometer centre and not at the test mass unique centre.

The selection of a 6–18 hour helio-synchronous orbit leads to the avoidance of fluctuations of the Sun orientation and the eclipse during the mission measurement phase, reducing the temperature variations and the thermo-elastic structural constraints which can generate spikes of acceleration. The same orbit is selected for GOCE but at lower altitude to the benefit of the gradiometer resolution but so with unfortunately many eclipses. The eccentricity of the quasi circular orbit must be sufficiently small to concentrate the power spectrum of the Earth monopole gravity field in one major spectral line. Furthermore, the fluctuations of the gravity gradient in the instrument frame has to be confined to the same spectral line.

The perturbing effect of the satellite and the Earth gravity gradient is of peculiar importance. The cylindrical test masses present sphere-like inertia matrices to limit these effects and the off-centering of the two masses is limited by construction to $20 \mu\text{m}$; gravity attractions are then equal on both masses with an accuracy of better than one part per thousand, which is sufficient for the experiment according to the environment. Moreover, the in orbital plane off-centering can be evaluated in flight with an accuracy better than $0.1 \mu\text{m}$ through the measured effect of the gravity gradient major component at twice the frequency of the gravity field.

Thus the orbit eccentricity must be smaller than 5×10^{-3} to limit at f_{EP} the disturbing difference of accelerations induced by the gravity gradients (in case of inertial pointing satellite; for a rotating satellite, the disturbing effect is modulated and so no more at f_{EP}). In this process, the knowledge of the satellite position is necessary to compute the gravity gradient: 300 m can be achieved along the three directions and appears sufficient. The out of plane off-centering cannot be estimated in orbit through the gravity gradient data analysis, thus the $20 \mu\text{m}$ initial value has to be considered when specifying the instrument axis direction in the orbital plane. Instrument frame, orbital plane and satellite rotation axis shall thus be aligned with a few 10^{-3} rd accuracy.

All these mission requirements correspond to the objective of 10^{-15} EP test accuracy, *i.e.* the detection at f_{EP} and in phase of any signal as weak as $8 \times 10^{-15} \text{ ms}^{-2}$, an integration period of 10^5 s being considered.

5. GOCE Gradiometry Mission

The GOCE (Gravity field and steady-state Ocean Circulation Explorer) mission was selected in 1999 as the first core mission of the Earth Explorer programme of ESA (European Space Agency). After numerous studies for more than two decades (Bernard *et al.*, 1983; Rummel and Colombo, 1985; Touboul *et al.*, 1991; Balmino

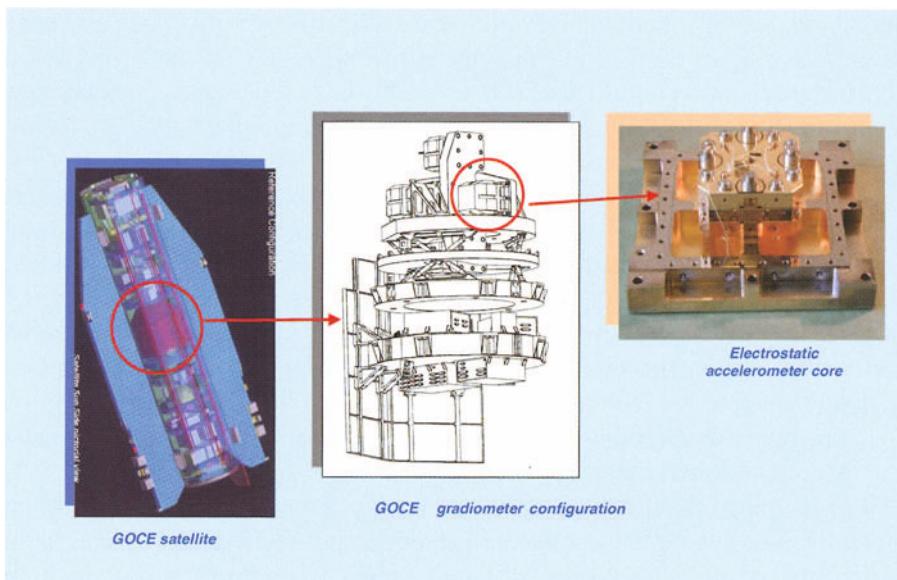


Figure 3. GOCE payload configuration. The satellite carries at its centre the gradiometer case with two thermally controlled housings. The gradiometer is composed of a rigid diamond structure, 6 inertial sensor cores and 4 electronics units, one per gradiometer axis plus a main interface unit.

and Perosanz, 1997), GOCE will fly for the first time a gradiometer on board a drag-free satellite at a very low altitude around 250 km (see Figure 3). The orbit inclination is 96.5° corresponding to an heliosynchronous orbit, 6h–18h for full and steady solar pointing. In fact, the satellite will be injected at a safer altitude of 270 km where the operations of all sub-systems and in particular the drag compensation will be verified before decreasing the altitude to 250 km and 240 km for two phases of measurements. The satellite cross section is minimum to reduce the drag to be compensated and the solar panels are rigidly mounted on the satellite like for the MICROSCOPE satellite.

The gradiometer instrument, designed by ONERA in the last decade, includes six up-to-date ultra-sensitive electrostatic accelerometers capable of measuring accelerations as weak as $4 \times 10^{-13} \text{ ms}^{-2}$ with 10 s integrating time. The gradiometer will deliver the three diagonal components of the gravity gradient tensor, T , in the Earth pointing spacecraft axes, with an accuracy of a few milli-Eötvös/Hz $^{1/2}$ (1 Eötvös = 10^{-9} s^{-2}), the frequency measurement bandwidth being from 0.005 Hz up to 0.1 Hz. The satellite is also tracked by the GPS satellite net and by ground laser stations, actually like CHAMP and GRACE. The fine trajectory recovery associated to the gradiometer measures is expected to conduct to a global gravity model with a resolution of at least 100 km and a total uncertainty of 2.5 millimeters on the geoid and 0.1 milligal on the gravity. This will therefore be an outstanding accurate reference for the concerned disciplines: geodesy, solid Earth physics,

oceanography and climatology. GOCE is scheduled for a launch at beginning of 2006.

The GRADIO accelerometer devoted to space gradiometry is in fact directly derived from the STAR and SuperSTAR accelerometers of respectively the CHAMP and the GRACE missions with quite the same mechanical core configuration and electrical architecture, which is well suited to realise a three-axes accelerometer with the possibility to perform ground tests under normal gravity. The parallelepiped proof-mass of 4 cm side and 1 cm height is made in Platinum-Rhodium instead of Titanium increasing the mass from 72 grams to 320 grams to the benefit of the rejection of the parasitic forces applied directly on the mass independently of the electrostatic suspension. The same technology as for the previous missions and for MICROSCOPE are implemented with gold coated ULE parts. The asymmetry of the configuration (not a cubic proof-mass) is required to perform the electrostatic suspension of the mass under one g, the electrode areas being larger along the vertical and the distance between the proof-mass and the electrodes 10 times less than in the two other directions, *i.e.* 30 μm . An electrostatic field as strong as $3 \times 10^7 \text{ V/m}$ is necessary to sustain the proof-mass. This axis will have in flight less performance than the two others (about a factor 300) with a greater full range and a stiffer mass control.

The electronics is made on the basis of the already developed circuits for the capacitive sensing and for the generation of the suspension control voltages. Optimal settings are defined according to the needed maximum range of operation that is forced by the DC gravity gradient to be sustained while the drag-free point is at the centre of the gradiometer. Contrarily to MICROSCOPE, this range is necessarily greater than a few 10^{-6} ms^{-2} and the stability of the accelerometer scale factor at level of 10^{-7} is mandatory for the two ultra-sensitive axes.

The thermal stability of the gradiometer case is thus demanded not only to preserve the stability of the sensor sensitivity and bias but also the geometry of the structure and the mounting parts. Table I compares the requirements with respect to the MICROSCOPE ones. While the GOCE requirements are more stringent in term of random fluctuations because of the huge signal to noise ratio to be reached, more than 10^6 even for the difference of the two measures of acceleration (from the DC gravity gradient of a few thousand Eötvös to the objective in milli-Eötvös per $\text{Hz}^{-1/2}$), the MICROSCOPE experiment demands also a peculiar care on the temperature fluctuations at the experiment frequency and in phase with the eventual EP violation signal.

6. GOCE and MICROSCOPE Similar Driving Parameters

The expression of the measure provided by each one axis gradiometer is identical to equation (1) but the Eötvös ratio (gravitational over inertial mass) is assumed

one:

$$\begin{aligned}
 \frac{\hat{F}_A}{m} - \frac{\hat{F}_B}{m} &\approx + (K_A - K_B) \frac{\ddot{X}_A + \ddot{x}_A + \ddot{X}_B + \ddot{x}_B}{2} + \left(I + \frac{K_A + K_B}{2} \right) \\
 &\quad \times (\ddot{X}_A - \ddot{X}_B) + \left(I + \frac{K_A + K_B}{2} \right) \\
 &\quad \times \{ (\ddot{x}_A - \ddot{x}_B) + 2\Omega (\dot{x}_A - \dot{x}_B) + (\Omega\Omega + \dot{\Omega})(x_A - x_B) \} \\
 &- \frac{\hat{F}_{p_A}}{m} + \frac{\hat{F}_{p_B}}{m} + \frac{E(F_A)}{m} - \frac{E(F_B)}{m} + \frac{E_{n_A}}{m} - \frac{E_{n_B}}{m} \\
 &- (g_A - g_B)
 \end{aligned} \tag{2}$$

The expected resolution of the accelerometer is $10^{-12} \text{ ms}^{-2}/\text{Hz}^{-1/2}$, along the gradiometer axis and the second ultra-sensitive axes, in the gradiometer measurement bandwidth of ($5 \times 10^{-3} \text{ Hz}$ – 0.1 Hz) corresponding to about 1400 km – 70 km track distance. As for the MICROSCOPE inertial sensor configuration, the resolution is limited at low frequency by the damping induced by the gold wire used for the mass charge control. In the upper frequency band, the limitation of the resolution is the noise, x_{noise} , of the proof-mass capacitive sensing: this acceleration noise source increases to the power 4 of the signal frequency f_{sig} (in PSD).

$$\Gamma_{\text{posnoise}}^2 = x_{\text{noise}}^2 \left(4\pi^2 f_{\text{sig}}^2 + 4\pi^2 f_p^2 \right)^2 (\text{ms}^{-2})^2/\text{Hz}$$

with f_p the frequency of the residual stiffness between the mass and the instrument frame, managed lower than f_{sig} by the selection of the electrostatic configuration and the diameter of the gold wire ($5 \mu\text{m}$).

In fact, in spite of the difference of the geometrical configuration, the intrinsic performance of the inertial sensors are very similar along the ultra-sensitive axes.

In GOCE, as in MICROSCOPE, all accelerometer outputs are also used for the drag compensation and the attitude control of the satellite. Because of the limited matching of the sensitivity matrices, both missions are very demanding for this satellite system. Table II summarises the established requirements in line with the mission objectives. Obviously, the frequency aspects are different for the two missions and MICROSCOPE is only demanding for the rejection of the disturbing line at the EP test frequency.

The requirement for the satellite pointing and positioning comes from the difference of the expression of the Earth gravity gradient tensor in the actual or the evaluated frame of measurement. Because in MICROSCOPE, this disturbing signal is limited by the mass centering ($20 \mu\text{m}$ instead of 0.5 m), the requirements are less stringent. The sensitivity matrices of the MICROSCOPE sensors are expected to be matched in orbit with a 3×10^{-4} relative accuracy and a 10^{-5} accuracy for the GOCE mission. The calibration process has been demonstrated in laboratory as shown in Figure 4. This more accurate calibration is necessary but also possible

TABLE II

Satellite controls. Both satellites are drag free thanks to Field Effect Electrical Propulsion and Ion thrusters. GOCE lower altitude (250 km, three times lower) requires much more control strength.

	MICROSCOPE	GOCE	GOCE
Satellite position	300 m	0.5 m	0.5 m
Satellite pointing	10^{-3} rd; 3×10^{-3} rdHz $^{-1/2}$	7×10^{-4} rd; 2×10^{-5} rdHz $^{-1/2}$	7×10^{-4} rd; 2×10^{-5} rdHz $^{-1/2}$
Satellite residual drag	3×10^{-10} ms $^{-2}$ Hz $^{-1/2}$	2.5×10^{-8} ms $^{-2}$ Hz $^{-1/2}$	2.5×10^{-8} ms $^{-2}$ Hz $^{-1/2}$
Angular velocity:			
Inertial pointing or		Earth pointing: 1.2×10^{-3} rds $^{-1}$	10^{-5} rds $^{-1}$
PSD	Rotating at 10^{-3} rds $^{-1}$	10^{-6} rds $^{-1}$ Hz $^{-1/2}$	5.10^{-7} rds $^{-1}$ Hz $^{-1/2}$
		(about $f_{EP} \sim 10^{-3}$ Hz)	(5×10^{-3} Hz-0.1Hz)
Tone (sine at fEP)	2×10^{-8} rds $^{-1}$	Not applicable	Not applicable
Angular acceleration:			
PSD	10^{-8} rds $^{-2}$ Hz $^{-1/2}$	10^{-6} rds $^{-2}$ Hz $^{-1/2}$	$1.5 \cdot 10^{-8}$ rds $^{-2}$ Hz $^{-1/2}$
	(about $f_{EP} \sim 10^{-3}$ Hz)		
Tone (sine at f_{EP})	2×10^{-11} rds $^{-2}$	Not applicable	Not applicable

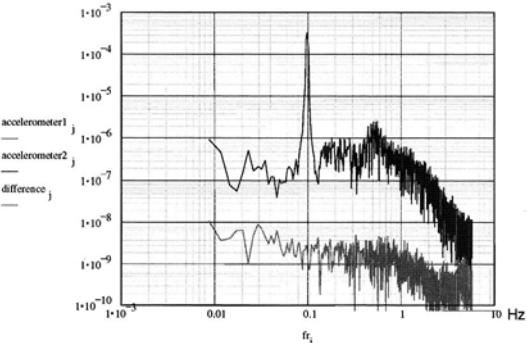
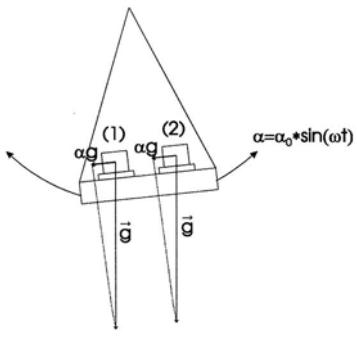


Figure 4. Laboratory calibration of the one axis gradiometer: along the gradiometer axis, the two inertial sensors are submitted to the one g gravity, projected according to the pendulum bench oscillation (α). One of the two sensor outputs is used for the servo-control of the orientation of the pendulum. A sine wave calibrated signal offsets the loop at 0.1 Hz and produces the oscillations. The difference of the two accelerometer outputs (in grey) demonstrates the equality of the sensitivities with a 10^{-5} accuracy and the rejection of the common mode signal (in black).

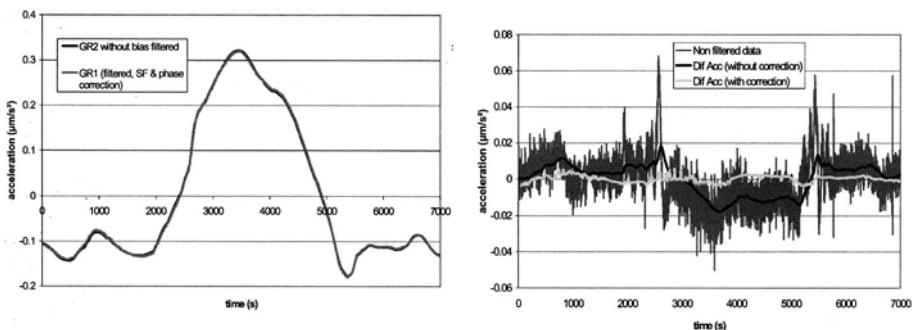


Figure 5. Data provided along track by the two accelerometers on board the GRACE satellites flying on the same orbit at 480 km altitude with 200 km separation. The sensitivities are set with a 4×10^{-3} relative accuracy. The difference of the two outputs are drawn in the right picture, before (in grey) and after (in black and white) filtering. The black and white curves show the adjustment of the sensitivities.

because of larger calibrating signals induced by stronger thrusters (needed however by the low altitude drag) and sustainable by the sensors. The MICROSCOPE satellite is also a 150 kg micro-satellite and procedures and software must be more simple.

The residual drag depends not only on the gain of the drag compensation system loop defining the filtering efficiency but also on the initial level of the atmospheric drag, more than one hundred larger at GOCE orbit altitude. The attitude control is similar in both cases.

7. Conclusion

In spite of the fully different scientific objectives of the GOCE mission dedicated to the global and accurate determination of the Earth gravity field and of the MICROSCOPE mission devoted to the test of the Equivalence Principle, these two missions are presently developed on the basis of the same type of electrostatic inertial sensors. Furthermore, the demanded performance of the instruments are very similar and, besides the difference of the geometry of the core, the operation, the involved technology and the limitations are really comparable. Both missions require a satellite with a drag compensation system and a very fine attitude control but the very low altitude of GOCE, as low as 240 km, is much more demanding for the electrical propulsion and the overall sub-system. The accommodation of the sensors needs in both mission a very stiff and steady structure in an insulated thermal case: an active control is envisaged on the GOCE satellite, a fully passive one appears sufficient for MICROSCOPE taking advantage of the weak instrument power dissipation, the Sun synchronous 6–18 hour orbit, the double stage insulation with an external baffled radiator always towards anti-Sun.

The specific difficulty of the GOCE mission is the huge range of measurement, a few 10^{-6} ms^{-2} (!), to be met relatively to the very high required resolution of $10^{-12} \text{ ms}^{-2} \text{ Hz}^{-1/2}$ in the bandwidth from $5 \times 10^{-3} \text{ Hz}$ to 10^{-1} Hz . The MICROSCOPE mission appears therefore easier to be performed with instrument operation at null. Nevertheless, the 10^{-15} EP test accuracy relies in the possibility to integrate during 10^5 seconds periods the measured signal in order to reach the detection of acceleration as weak as $3 \times 10^{-15} \text{ ms}^{-2}$ with the sensor resolution limited to $10^{-12} \text{ ms}^{-2} \text{ Hz}^{-1/2}$.

At last, in flight data from the present CHAMP and GRACE satellites show that accelerometers on board satellite are very sensitive to the satellite thermo-elastic behaviour and motion. Figure 5 provides an example of atmospheric drag measurements (along track) by both accelerometers on board the quite identical GRACE satellites flying on the same orbit.

The two curves (left) are very similar and show that the atmospheric density does not vary during the 30 s period between the two satellite passes. The difference of the two curves is drawn with a larger scale (right) showing much more signal oscillations at higher frequency. Clearly, the success of both the MICROSCOPE and the GOCE mission will rely on the soft environment managed on board the satellite and on the filtering ratio that will be obtained in the measurement frequency bandwidth through the differential signal.

References

- Balmino, G. and Perosanz, F.: 1997, 'Comparison of geopotential recovery capabilities of some future satellite missions', *Proc. IAG symposium, Gravity and Geoid*, p. 403–412.

- Bernard, A., Sacleux, B., and Touboul, P.: 1983, 'GRADIO: orbital gravity gradiometry through differential micro accelerometry', 34th IAF Congress, Budapest (Hungary), October 1983, ONERA TP 83 – 18.
- CNES: 2002, 'The French scientific space program', CNES Report to 34th COSPAR 2002, CNES publication.
- Damour, T., Piazza, F., and Veneziano, G.: 2002, 'Violations of the equivalence principle in a dilaton-runaway scenario', IHES/P/02/09, Bicocca-FT-02-03, CERN-TH/2002-093, May 2002.
- ESA: 1999, 'Gravity field and steady state Ocean Circulation Explorer mission', ESA SP-1233 (1).
- Fischbach, E. and Talmadge, C. L.: 1998, *The Search for non Newtonian Gravity*, Springer Verlag, 320 pp.
- JPL, GRACE: 1998, 'Gravity Recovery and Climate Experiment — Science and Mission Requirements Document', rev. A, JPLD-15928, NASA's Earth System Science Pathfinder Program 1–84.
- Josselin, V., Touboul, P., and Kielbasa, R.: 1999, 'Capacitive detection scheme for space accelerometers applications Sensors and Actuators A', *Physical.* **78**, 92–98.
- Lämmerzahl, C., Everitt, C. W. F., and Hehl, F. W. (eds.): 2000, *Gyros, Clocks, Interferometers...: Testing Relativistic Gravity in Space, Lecture Notes in Physics*, Springer Verlag, 507 pp.
- McPherson, K. M., et al.: 1999, 'A Summary of the Quasi-Steady Acceleration Environment On-Board STS-94 (MSL-1)', AIAA, Reno, USA.
- Nati, M. et al.: 1994, 'ASTRE, a highly performant accelerometer for the low frequency range of the microgravity environment', 24th Symposium on space environmental control systems, Friedrichshafen, Germany, AIAA Proceedings.
- Reigber, Ch., et al.: 1996, CHAMP Phase B Executive Summary, G.F.Z., STR96/13 1–37.
- Reigber, Ch., Lühr, H., Schwintzer, P.: 2002, 'Champ mission status', *Advances in Space Research* **30**(2), 192–134.
- Rummel, R. and Colombo, O. L.: 1985, 'Gravity field determination from satellite gradiometry', *Bull. Geod.* **59**, p. 233–246.
- Tapley, B. and Reigber, C.: 2002, 'The GRACE mision, its status and early results', 34th COSPAR, Houston, October 2002.
- Touboul, P.: 2000, 'Space accelerometers: present status', in C. Lämmerzahl, C. W. F. Everitt, F. W. Hehl (eds.), *Gyrod, Clocks, Interferometers...: Testing Relativistic Gravity in Space, Lecture Note in Physics*, Springer Verlag, p. 274.
- Touboul, P. et al.: 1991, 'Drag effect on gradiometer measurements', *Manuscripta Geodaetica* **16**(2), 73–91.
- Touboul, P., Foulon, B., and Le Clerc, G.: 1998, 'STAR the accelerometer of the geodesic mission CHAMP', IAF-98-B.3.07.
- Touboul, P., Foulon, B., and Lafargue, L.: 2002, 'The MICROSCOPE mission', *Acta Astronautica* **50**(7), 433–443.
- Willemenot, E. and Touboul, P., 'On-ground investigations of space accelerometers noise with an electrostatic torsion pendulum', *Review of Scientific Instruments* **71**/**1**, 302–309.

NEEDS AND TOOLS FOR FUTURE GRAVITY MEASURING MISSIONS

MIGUEL AGUIRRE-MARTINEZ (miguel.aguirre@esa.int)
European Space Agency, European Space Technology and Engineering Centre

NICO SNEEUW (sneeuw@ucalgary.ca)
Department of Geomatics Engineering, University of Calgary

Received: 17 May 2002; Accepted in final form: 4 July 2002

Abstract. This paper compares the requirements that can be expected of gravity measuring missions with respect to the status of the instrumentation and satellite technologies. The error sources of gravity gradiometry and satellite-to-satellite tracking are analysed and the elements limiting the accuracy are identified. Proposed and approved future missions that will fly technologies of interest for gravity sensing are recalled. Areas of technical development of interest are reviewed. The article finishes with two possible conceptual missions presented as examples and with a chapter of conclusions.

Keywords: gravity field, satellite missions, technology developments, GRACE, GOCE

1. Background

After implementation of GOCE and GRACE, work on the architecture of new missions is already starting. The shape of the future missions will depend on the—often conflicting—demands of the different scientific domains and the status of maturity of the different technologies.

2. The Error Structure of Gravity Measuring Missions

Gravity field satellite missions measure gravity related observables with a given level of accuracy, with respect to a given spatial scale. Satellites rotate around the Earth. Therefore, the spatial scale becomes a frequency bandwidth requirement and the accuracy becomes a requirement integrated over the specified bandwidth. The error structure of GOCE can be seen in (ESA, 1999, ch. 6 and 8). The total mission accuracy target is $4 \text{ mE}/\sqrt{\text{Hz}}$ in the bandwidth 0.005–0.1 Hz. In the case of GRACE the mission accuracy shall be around $10^{-2} \mu\text{m}/(\text{s}^{-2}\sqrt{\text{Hz}})$ in the range 0.0001–0.02 Hz, taking into account the 3 noise sources (accelerometer noise, USO stability and microwave system noise), cf. (Davis et al., 1999, fig. 6). The accuracy will define the required level of quality and the frequency range will define the time-scale where this quality should be ensured.

From the mission design point of view, gravity-sensing errors can be organised in three groups: Instrument, Satellite and Coupled. The instrument errors depend



only on the gravity-sensing instrument. The most important is the instrument intrinsic noise. The coupling errors depend on the performances of the instrument and the spacecraft together, e.g. coupling of the satellite external perturbations with the misalignments of the instrument. The satellite errors depend only on the satellite. In this category we can include attitude recovery and pointing errors and errors in the determination of the external accelerations.

The instrument is only one of the possible sources of mission errors and the overall performance of a gravity-sensing mission will depend strongly on the adequate matching of the instrument with the satellite platform carrying it. A properly designed mission will balance the difficulties of the instrument and the satellite. This needs for balance is acknowledged in (Davis et al., 1999, p. 9).

3. The Limitations of GRACE and GOCE

A detailed error breakdown of GOCE can be seen in (ESA, 1999, tbl. 6.1). The most important error source is the intrinsic noise of the gradiometer that takes 3 of the 4 mE of total mission error allocation. Therefore, the sophisticated GOCE satellite attitude and drag control elements would be compatible with an even more accurate instrument.

The most comprehensive analysis of the error structure of GRACE can be seen in (Tapley, 1996, figs. 3–8). Two of them are dominant at high frequency: the microwave system noise and the accelerometers' attitude and location errors. At low frequencies the accelerometers' intrinsic noise and the temperature control of the base-plate are the dominant error sources. GRACE is equally limited by the quality of the satellite and the instrument.

On top of the performances of the instrument and the satellite, the mission performance is going to be limited by the flying altitude and the mission duration. GOCE uses electrical propulsion to null the effect of the atmospheric drag. The nominal mission is 225 days at 250 km and 225 days at 240 km. GRACE has no altitude control. Mission duration can be up to 1700 days decaying from 500 to 300 km. The use of the orbit control tools of GOCE will make future GRACE-like missions more interesting.

Future missions should overcome the aforementioned limitations. A future GOCE-type mission requires a better instrument, whereas a post-GRACE mission needs a better satellite. In the case of a future GRACE-like mission, these attitude and drag control improvements are even more important if optical tracking is used. Lower altitudes and longer duration would be necessary for both mission types. Again, a post-GRACE mission will improve significantly by using a more sophisticated satellite which allows flying at constant low altitude for long periods of time.

4. New Needs and Proposed Missions

LISA is a constellation of satellites that track each other with high accuracy. It will study gravitational waves produced by galactic and extragalactic binaries (neutron stars, white dwarfs, black holes). According to (LISA, 2000), the accuracy target is a sensitivity of $20 \text{ pm}/\sqrt{\text{Hz}}$ at 1 mHz. The measurements frequency range is 10^{-4} to 0.1 Hz. The attitude and drag free performance targets of LISA are 1000 times more demanding than for GOCE, but the technologies used to provide this performance are very similar. This can be explained because the deep space environment of LISA is much more benign than the low Earth orbit environment of GOCE.

If translated directly to gravity sensing performances, the position sensing accuracy of LISA would allow the implementation of missions to study the dynamic of the Earth liquid core. LICODY is an example of this type of mission. According to (Greft-Lefftz et al., 2000), it requires an accuracy of around 5 pm/s velocity disturbance detection—5 orders of magnitude better than GRACE—in the frequency ranges: 0–0.001 Hz (core studies) and 0.001–0.5 Hz (solid Earth and climate).

Nevertheless these targets will not be met unless implemented on a system where the external perturbations are reduced as much as needed to arrive at the expected performance. As said before, the attitude and position control tools of LISA are similar to the tools used by GOCE. Tools able to provide LISA-like attitude and position control performances in an orbit around the Earth are not yet on the horizon.

5. The Advantages and Problems of Low Altitude Flying

According to Lambeck (1988) the strength of the gravity signal associated with the spherical harmonic degree n at an orbit altitude of h is proportional to $(R/(R+h))^n$. The exponent n implies that the lower the mission altitude, the higher the resolution of the gravity field recovery will be. Low altitude improves the performance of any mission. The price to pay for the enhanced gravity signal is a strong atmospheric drag.

GOCE flies at 250 km altitude and has a cross-sectional area of 0.8 m^2 . The shape of the Earth produces large changes in the value of the drag along the orbit, the maximum drag force is around 10 mN and the minimum around 2 mN. To compensate this force ion propulsion is used. The amount of power required is reasonable: 475 W.

The drag dominant frequencies are the orbital one and its harmonics. They can be clearly seen in Fig. 1. The higher the frequency, the lower the power of the drag. The decrease of power with the frequency is roughly proportional to $1/f$.

A future gradiometric mission is likely to require a cryogenic gradiometer and it will be very difficult to reduce the present GOCE transversal area of 0.8 m^2 . On

Noise power spectral density of the drag acceleration

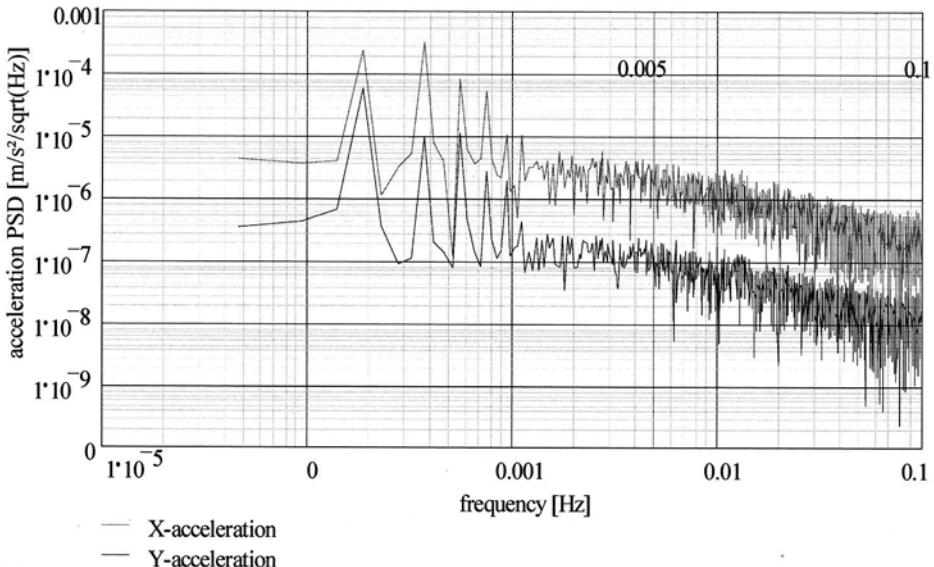


Figure 1. The frequency behaviour of the atmospheric forces acting on GOCE at 250 km altitude.

the other hand, GRACE-like missions with laser interferometry have the potential for miniaturisation. In this case the transversal area could be greatly reduced.

6. Implementation of Drag Control in Low Earth Orbit

The drag-free satellite was originally proposed by Lange (1962). GOCE is demonstrating that drag free control in low Earth orbit is possible. As sensors it carries a star imager and the gradiometer, acting as a gyroscope. Moreover it uses two types of ion thrusters, cf. Fig. 2, for orbit maintenance and drag control. In the measurement bandwidth, GOCE requires a control better than $9 \mu\text{rad}/\sqrt{\text{Hz}}$ in attitude and $2.5^{-8} \text{ m}/(\text{s}^2 \sqrt{\text{Hz}})$ in acceleration.

7. Possible Future Missions

The two mission scenarios described in this section were analysed by a team of graduate students during a summer course under the direction of the authors. The study included a rough verification of the feasibility of the mission parameters.

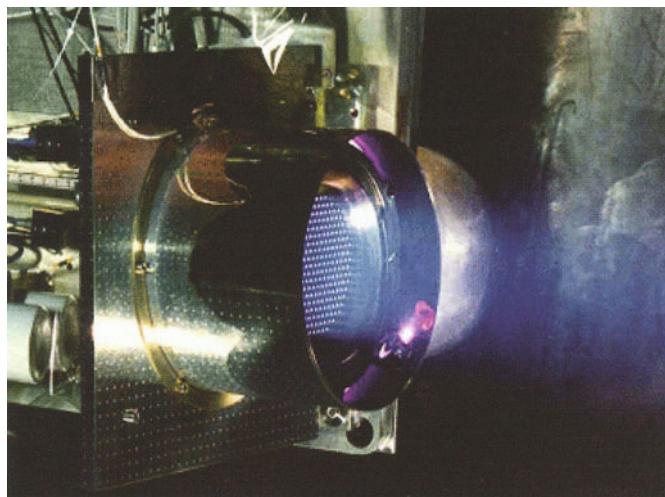


Figure 2. GOCE ion thruster testing at ESTEC.

7.1. FUTURE IMPROVED GRAVITY GRADIOMETRY MISSION

The accuracy target for this mission scenarios was established to be $0.7 \text{ mE}/\sqrt{\text{Hz}}$. The error budget was broken-down into $0.5 \text{ mE}/\sqrt{\text{Hz}}$ for the instrument, $0.3 \text{ mE}/\sqrt{\text{Hz}}$ for the errors depending on the external accelerations and $0.3 \text{ mE}/\sqrt{\text{Hz}}$ for the recovery of the angular accelerations. For this mission the attitude and drag control architecture can be identical to GOCE but the power for the ion thrusters shall increase from 475 to 700 W. This will allow the doubling of the maximum drag the satellite is able to control (from 10 to 20 mN). The resulting overall satellite power demand will be around 1300 W.

As in GOCE, the satellite mass target would be 1000 kg. The configuration will be almost identical to GOCE. Dnieper was chosen as the reference launcher. The orbit selected was sun-synchronous at 210 km of altitude and the mission duration will be 24 months.

7.2. FUTURE SATELLITE-TO-SATELLITE TRACKING MISSION

In this mission scenario two satellites perform satellite-to-satellite tracking by optical interferometry. They are separated by 150 km. The accuracy target for the whole mission will be $0.07 \mu\text{m}/(\text{s} \sqrt{\text{Hz}})$. In first approximation and assuming a laser link between both satellites, the satellite will need to be pointed with an accuracy of 0.1 mrad and the optics should have a fine pointing mechanism with an accuracy of $2.5 \mu\text{rad}$. The instrument will need a dynamic range of 10 m/s. The satellite will need high accuracy recovery of the linear and angular accelerations at least down to the GOCE quality level. This will require the development of miniature angular and linear acceleration sensors.

The possibilities allowed by optics miniaturisation will allow a mass target of 150 kg for each satellite. This will make the pair compatible with the smallest launchers like Start-1 or Pegasus. The target transversal area will be around 0.4 m².

The satellite will have drag free capability provided by a single FEEP type ion propulsion with a slit width of 50 mm. This arrangement will provide up to 0.5 mN of thrust. The power consumed by this arrangement will be just 30 W. The reference orbit parameters were 90° of inclination and 310 km of altitude, the mission duration 48 months.

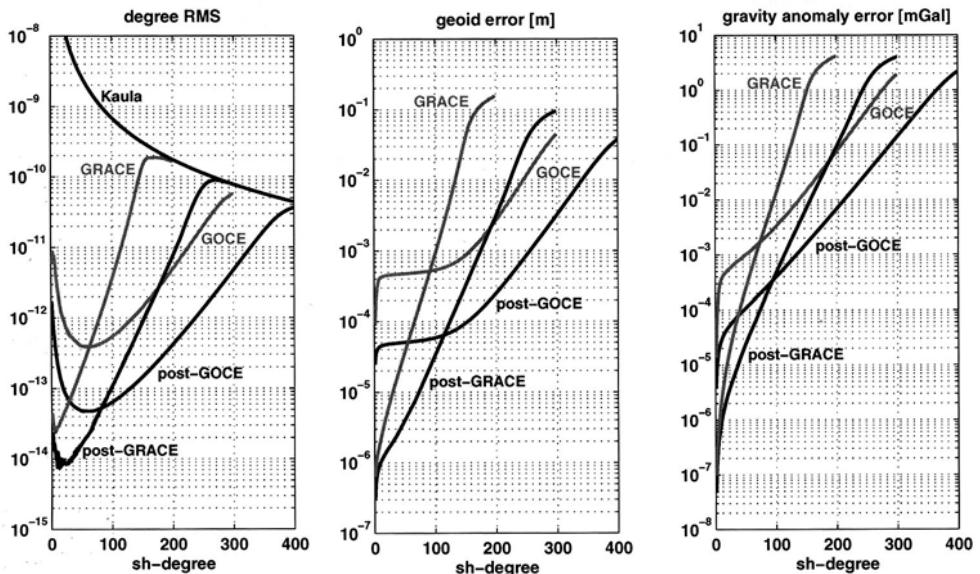


Figure 3. Performances of proposed post-GRACE and post-GOCE missions in term of degree RMS, cumulative geoid and cumulative gravity anomaly.

8. Mission Performances

The performances for the two reference missions described above have been analysed using the method described in (Sneeuw et al., 1996) and (Oberndorfer et al., 1999). The reference missions provide outstanding performance improvements over GRACE and GOCE by more than one order of magnitude.

Figure 3 provides the individual and cumulative errors as a function of the spherical harmonic degree. Figures 4 and 5 provide the number of “correct” significant digits in the determination of the spherical harmonic coefficients corresponding to the indicated degree and order.

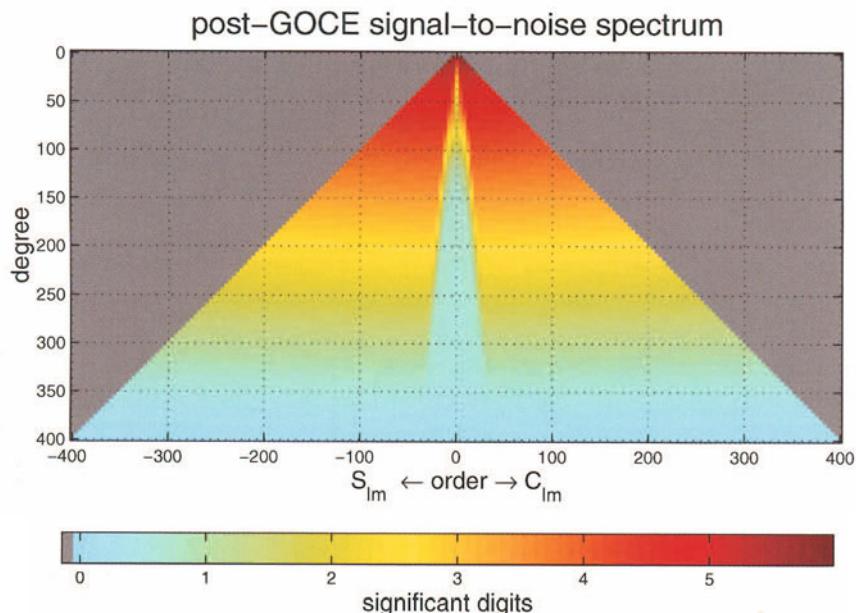


Figure 4. Post-GOCE spherical harmonic error spectra in terms of significant digits.

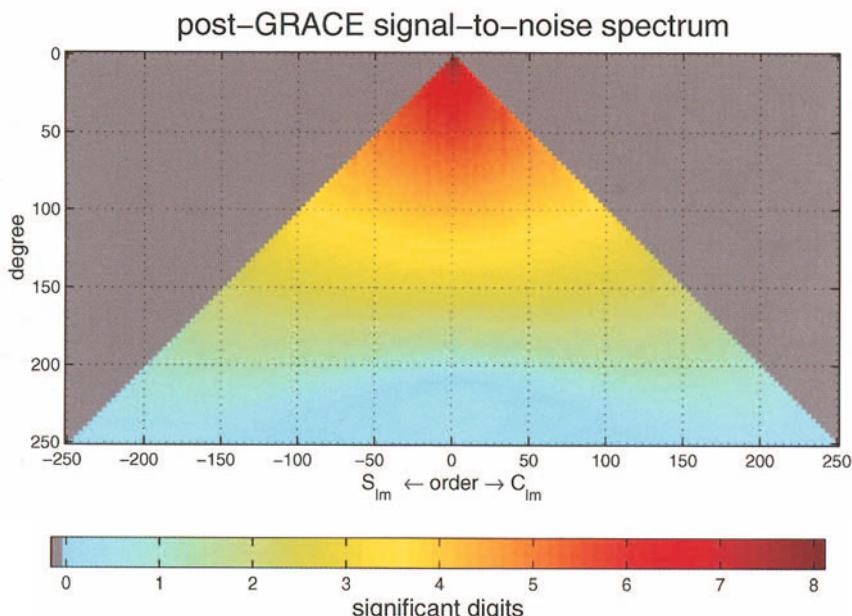


Figure 5. Post-GRACE spherical harmonic error spectra in terms of significant digits. Please note the different degree, order and colour scales for both reference missions.

9. Conclusions

- Cryogenic gradiometry should continue to be developed.
- Optical interferometry is the logical next step for low-low satellite-to-satellite tracking. Nevertheless, the requirements on the satellite driven by the use of this technology need to be studied in detail. Limitations of the satellite could make it very difficult to materialize the very promising performances of laser satellite-to-satellite tracking.
- Space science is developing technologies adequate for gravity sensing missions. However, the applicability of these technical solutions to low Earth orbit remains to be seen.
- Good instruments deserve good satellites.
- Orbit maintenance by electric propulsion will dramatically increase the performance of a post-GRACE mission with very little additional cost.

References

- Davis, E., et al.: 1999, 'The GRACE Mission: Meeting the Technical Challenges', IAF-99-B.2.05, *Proceedings of the 50th IAF Congress*.
- ESA: 1999, 'Gravity Field and Steady-State Ocean Circulation Mission', ESA SP-1233(1), *Report for mission selection of the four candidate earth explorer missions*, European Space Agency.
- Greft-Lefftz, M., et al.: 2000, LICODY, *Proposal for ESA core explorer selection*.
- Lambeck, K.: 1988, *Geophysical Geodesy*, Oxford Science Publications.
- Lange, B.: 1962, 'The drag free satellite', *AIAA Journal* **2**(9).
- LISA team: 2000, *LISA System and Technology Study Report*, ESA-SCI(2000) (11), European Space Agency.
- Oberndorfer, H., Dorobantu, R., Gerlach, C., Müller, J., Rummel, R., Sneeuw, N., Koop, R., Visser, P., Hoyng, P., Selig, A., and Smit M.: 1999, 'GOCE sensor combination and error analysis', *Bollettino di Geofisica Teorica ed Applicata* **40**, 303–308.
- Sneeuw, N., Koop, R., and Schrama, E.: 1996, 'Global Gravity Field Error Determination for the STEP Geodesy co-experiment', *ESA WPP-115*, 358–367.
- Tapley, B.: 1996, *Gravity Recovery and Climate Experiment Proposal to NASA Pathfinder Program*. Address for Offprints: Miguel Aguirre-Martinez, European Space Agency, European Space Technology and Engineering Centre, Keplerlaan 2, 2200 AG Noordwijk, The Netherlands

VII: CLOSING SESSION

GOCE: ESA'S FIRST EARTH EXPLORER CORE MISSION

M. R. DRINKWATER, R. FLOBERGHAGEN, R. HAAGMANS, D. MUZI and
A. POPESCU

European Space Agency, ESTEC, 2200 AG Noordwijk, The Netherlands
(mark.drinkwater@esa.int)

Received: 10 June 2002; Accepted in final form: 18 November 2002

Abstract. This paper introduces the first ESA Core Earth Explorer mission, GOCE, in the context of ESA's Living Planet programme. GOCE will measure highly accurate, high spatial resolution differential accelerations in three dimensions along a well characterised orbit: the mission is planned for launch in early 2006. The mission objectives are to obtain gravity gradient data such that new global and regional models of the static Earth's gravity field and of the geoid can be deduced at length scales down to 100 km. These products will have broad application in the fields of geodesy, oceanography, solid-earth physics and glaciology.

1. Introduction

The European Space Agency's (ESA's) "Living Planet Programme" (ESA, 1998: <http://www.esa.int/livingplanet>) defines the ESA strategy and plans for satellite Earth Observation (EO) in the 21st century. It marks the beginning of an era in which European EO missions are smaller and more focussed than their predecessors (*e.g.* ERS-1, -2, and Envisat). This programme is user-driven in terms of addressing science and research community measurement requirements with the Earth Explorer series of missions, and applications and operational requirements with the Earth Watch line of missions (Figure 1). Moreover, the main objectives of the programme are to further develop our knowledge of the complex Earth system; to preserve the Earth and its environment and resources; and to provide information with which to more efficiently and effectively manage life on Earth.

ESA Earth Explorer missions are designed to address EO topics using a combination of new technology and new scientific techniques. There are two categories of Earth Explorer missions: 'Core' and 'Opportunity'. Core missions respond directly to specific areas of public concern and are selected through wide consultation with the scientific and research communities. These missions are ESA-led and must be realised with a financial ceiling of \sim 350 MEuro, with a new mission launched approximately every two years. Opportunity missions are less complex and quicker to implement, and where possible they should use smaller, low-cost satellites and be realised with a financial ceiling of \sim 100 MEuro. Opportunity missions are designed to respond to new evolving areas of research or areas of immediate en-



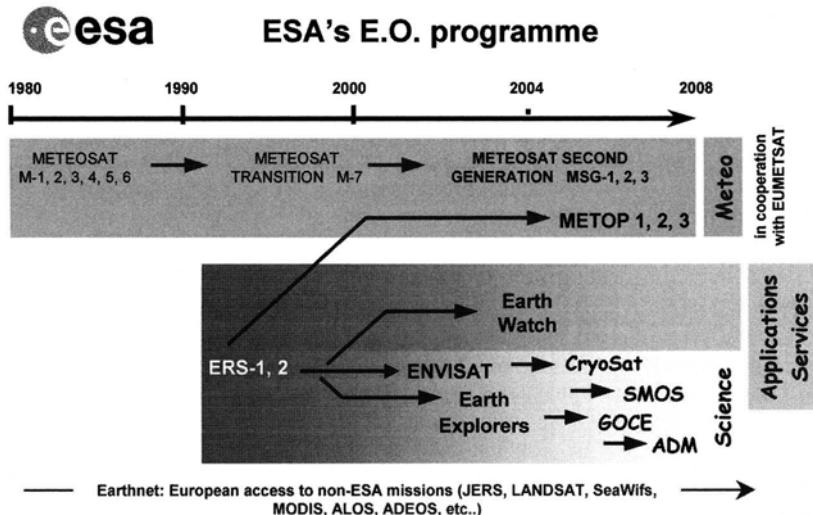


Figure 1. Evolution of the three main elements of ESA's Earth Observation programme. The lower bar indicates the development of two new parallel elements of the Living Planet Programme comprising the Earth Explorer series of missions and the Earth Watch line of future operational missions. Complementing these are the continuing meteorological satellite series and Earthnet.

vironmental concern. These missions can be developed with significant external contributions to ESA.

Out of nine Earth Explorer core missions proposed in the first Call for Core Explorers (*i.e.* Announcement of Opportunity) in 1996, two missions were ultimately selected for implementation in 1999. These were the Gravity field and steady-state Ocean Circulation Explorer (GOCE) and the Atmospheric Dynamics Mission (ADM-Aeolus), respectively. GOCE will be the first of these two missions, with launch scheduled in 2006 with a two-year mission duration (ESA, 2001).

2. GOCE Mission Objectives

The primary GOCE mission objectives are to:

- determine the Earth's gravity field with an accuracy of 1 mGal (where $1 \text{ mGal} = 10^{-5} \text{ m/s}^2$) via the measurement of the components of the gravity gradient tensor in combination with satellite to satellite tracking
- determine the geoid (*i.e.* equipotential surface for a hypothetical ocean at rest) with an accuracy of 1 cm
- achieve both of the above at length scales down to 100 km (*i.e.* spherical harmonic degree and order 200).

The GOCE mission serves to support the following multi-disciplinary science objectives:

- to permit precise estimation of the quasi-static marine geoid, needed for the quantitative determination, in combination with satellite altimetry and/or *in situ* data, of absolute ocean circulation and transport of mass.
- to provide a better understanding of the physics of the Earth's interior including geodynamics associated with the lithosphere, mantle composition and rheology, uplifting and subduction processes.
- to provide a high-accuracy global height reference system for datum connection. This may serve as a reference surface for the study of topographic processes, including the evolution of ice sheets and land surface topography; and in GPS Levelling applications.
- to estimate the mass and thickness of the polar ice sheets through a combination of bedrock topography, derived from gravity anomalies and ice-sheet surface elevation (from altimetry).

GOCE addresses two of the themes that underpin the ESA Living Planet Programme. These are Theme 1 — the Earth Interior, and Theme 2 — the Physical Climate System, respectively. GOCE addresses the first by making measurements to address the geoid, the gravity field at various scales, from local and regional to global (see Table I). The second theme is addressed by GOCE in terms of the importance of the derivation of the marine geoid for effective use of other satellite data (such as satellite altimetry) in calculation of the absolute ocean circulation.

3. GOCE Mission History

The *Gravity field and steady-state Ocean Circulation Explorer* (GOCE) mission concept was first proposed and considered at the first User Consultation Workshop held in Granada, Spain in May 1996 (ESA, 1996) along with eight other candidates. The measurement principles exploited by the GOCE mission have a long history (Wolff, 1969; Rummel, 1979) and the concept was conceived in large part in prior preparatory studies for the Solid Earth Science and Application Mission for Europe (*SESAME*) in the 1980's (ESA, 1986) and subsequently the *Aristoteles* mission concept (ESA, 1991). Upon completion of the 1996 Granada Workshop, four mission selection recommendations were made by the ESA Earth Science Advisory Committee (ESAC) from the nine candidates. The Earth Observation Programme Board (PB-EO) subsequently considered the ESAC recommendations and endorsed the selection of GOCE for further detailed study.

Following the 1996 PB-EO approval of GOCE, a Mission Advisory Group (MAG) was established to support the Agency with advice during pre-Phase A studies, and to oversee supporting scientific studies. The MAG was first tasked with establishing scientifically-driven performance requirements in the form of a mission requirements document. In July 1998, a Phase A design feasibility study

TABLE I

Measurement requirements in terms of geoid height and gravity anomaly accuracies (after Rebhan *et al.*, 2000).

Application	Accuracy		Spatial Resolution half wavelength (km)
	Geoid (cm)	Gravity (mGal)	
<i>Oceanography:</i>			
- Short scale	1-2		100 km
	0.2		200 km
- Basin scale	~0.1		1000 km
<i>Solid Earth:</i>			
- Lithosphere and upper-mantle density structure		1-2	100 km
- Continental lithosphere			
• Sedimentary basins		1-2	50-100 km
• Rifts		1-2	20-100 km
• Tectonic motions		1-2	100-500 km
• Seismic hazards		1.0	100 km
- Ocean lithosphere & interaction with asthenosphere		0.5-1.0	100-200 km
<i>Geodesy:</i>			
- Levelling by GPS	1.0		100-1000 km
- Unification of worldwide height systems	1.0		100-20000 km
- Inertial navigation system		~1-5	100-1000 km
- Orbit (1 cm radial orbit error for altimetric satellites		~1-3	100-1000 km
<i>Ice sheets:</i>			
- Rock basement		1-5	50-100 km
- Ice vertical movements	2.0		100-1000 km
<i>Sea-level change</i>			
	Many of the above applications, with their specific requirements, are relevant to sea-level studies		

was initiated with industry on the basis of the resulting system requirements. Upon conclusion of this study in July 1999, a final Report for Mission Selection (ESA, 1999) was drafted by the MAG and presented at the second User Consultation Workshop in Granada, Spain in October, 1999. At this workshop the four competing mission concepts were scientifically reviewed. Following this second Granada meeting GOCE was one of two Core Explorer missions to be recommended by ESAC. The PB-EO subsequently endorsed the ESAC recommendation during its November 1999 meeting and authorised the Executive's proposal to begin GOCE implementation as the first Core Explorer mission.

In April, 2000 the Invitation to Tender (ITT) for “Definition, Design, Development, Manufacture, Integration, Testing, Support to Launch and Commissioning of the Spacecraft, including Payload, for the GOCE mission” was released by ESA to industry. The GOCE Contract Proposal received from Alenia Spazio (of Italy) was accepted for review in July 2000 and a Tender Evaluation Board convened to carefully evaluate the proposal.

Approval of the GOCE Contract Proposal was required from the Industrial Policy Committee (IPC) to continue with Phase B/C/D/E1 activities. After successful negotiations with Alenia Spazio to revise and improve various details, the Executive invited the IPC to approve the placing of a contract with Alenia Spazio. Authorisation was finally given by the IPC in January 2001 for initiation of Phase B industrial design activities. The GOCE prime contractor, Alenia Spazio is supported by an industrial consortium including Astrium GmbH (of Germany), Alcatel Space Industries (of France) and ONERA (of France).

4. GOCE Current Status

4.1. INDUSTRIAL ACTIVITIES

Phase B GOCE satellite design activities were concluded in April 2002 with a thorough Space-Segment Preliminary Design Review (PDR), in which a consolidated, baseline spacecraft design was presented (Figure 2).

During the GOCE design phase (Phase B: December 2000–Feb. 2002), the industrial consortium members carried out the necessary analyses in order to refine and consolidate the satellite baseline design and the related spacecraft interfaces. Consolidation activities were a prerequisite to the preparation of equipment specifications included in various ITT packages. These ITT's were issued in 2001 as part of the competitive selection process of sub-contractors for the construction of various elements of the GOCE satellite, in line with the ESA industrial procurement policy.

A key technical investigation conducted during Phase B has been assessment of the robustness of the current accelerometer design along with its capability to withstand vibrations during the spacecraft launch. Each of the six individual accelerometer sensor heads within the Electrostatic Gravity Gradiometer (EGG) instrument has a proof mass approx. 4.6 times heavier than those in previous accelerometers developed and built by ONERA (such as SuperStar accelerometers on the two GRACE satellites). The intention of the heavier proof mass is to improve sensitivity, but in reality this also places much more stringent constraints on the GOCE spacecraft design and mission implementation.

Another area currently under careful investigation is the technical development of the GOCE spacecraft micro thrusters. Due to the need for precise attitude and orbit control, the micro-thrusters are subjected to extremely demanding requirements in terms of thrust performance, thrust quantisation, and noise power spectral

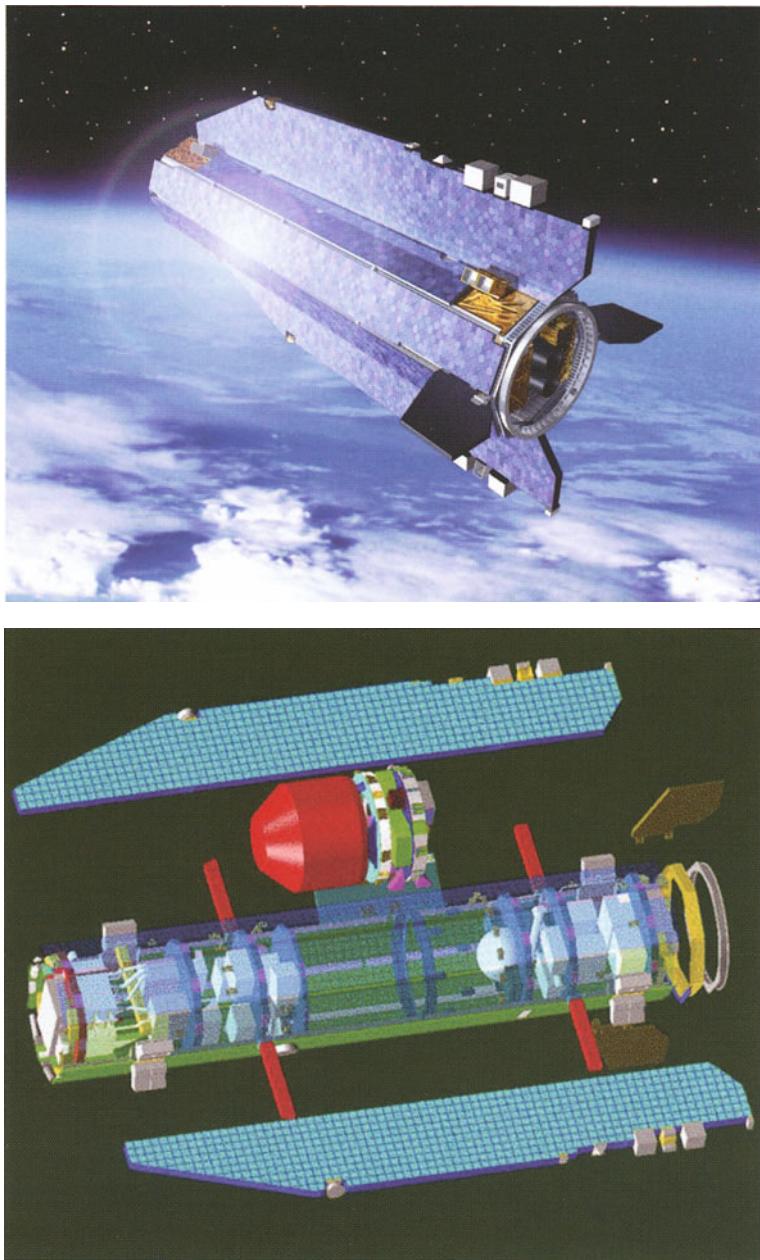


Figure 2. Upper panel: View of the sun-illuminated side of the GOCE spacecraft. *Lower panel:* view of the interior of the GOCE spacecraft with detached solar panels, winglets and the gradiometer assembly (courtesy Astrium GmbH and Alenia Spazio).

density (see Canuto *et al.*, 2003). After evaluation of the trade-offs for mass and micro-disturbances, field emission electric propulsion (FEEP) was selected for the micro-propulsion system rather than the cold gas option. This choice was made because of the FEEP technology's superior performance in relation to the minimisation of spacecraft-induced perturbations on the accelerometer sensor-heads. A final selection of the supplier will be made in the near future on the basis of the criteria related to performance, lifetime, cost and impact on the system.

An April 2002 PDR Board meeting considered the findings and conclusions of the design review, concentrating on what were considered to be residual problem areas. The Board reported positively that no technical "show-stopper" has been identified, confirming that the overall approach for the design of the GOCE Space Segment is robust. Nonetheless, some refinements to the present design are inevitable, and anticipated in order to comply with the challenging GOCE gradiometric performance requirements and spacecraft and instrument construction schedule.

4.2. FUTURE MISSION SCHEDULE

The positive outcome of the PDR signals the successful completion of Phase B, and the transition of GOCE project activities into the Construction Phase (Phase C). An approximate 3.5 year Phase C/D is envisaged, incorporating a Critical Design Review (CDR) in May 2004. A Flight Acceptance Review is planned for around November 2005, to be compatible with a launch date in 2006.

5. The GOCE Spacecraft

The GOCE mission employs the principle of gravity gradiometry. Satellite gradiometry requires measurement of acceleration differences between the test masses of an ensemble of accelerometers inside an orbiting vehicle. The measured signal is the difference in gravitational acceleration inside the spacecraft, where the gravitational signal reflects the pull of the Earth's varying gravity field caused by varying masses of mountains and valleys, ocean ridges and trenches, subduction zones and mantle inhomogeneities, etc. The measured signals correspond to the second derivatives of the gravitational potential. The gradiometer measurements are supplemented by high-accuracy Satellite-to Satellite Tracking (SST) measurements and star-tracker information.

Two core instruments are employed on GOCE: an Electrostatic Gravity Gradiometer (EGG) and an SST Instrument (SSTI). SSTI incorporates a Geodetic GPS receiver for high-low (hl) tracking between the satellites of the GPS constellation, and the low flying GOCE spacecraft (referred to as SST-hl). The EGG is a three-axis satellite gravity gradiometer, each arm of which comprises a pair of accelerometers. The gradiometer thus exploits the principle of differential accelerometry.

Drag-free attitude and orbital control, together with the common mode accelerations recorded by the pairs of accelerometers in the EGG, allow the separation of the gravitational signal from non-gravitational satellite skin forces and angular motion. Time variable effects of eigen-gravitation will be kept (by design) below the instrument noise level. The SSTI allows the retrieval of the long wavelength terms of the gravity field while the EGG is devoted to the medium and shorter wavelength terms. The gravity field information derived from both instruments overlap at low frequencies, around 0.005 Hz.

From the measurement principle point of view, the GOCE mission concept is unique in meeting four fundamental criteria for gravity field missions, namely:

- Uninterrupted tracking in three spatial dimensions
- Continuous compensation of the effect of non-gravitational forces
- Selection of a low orbital altitude for a strong gravity signal
- Counteracting of the gravity field attenuation at altitude by employing satellite gravity gradiometry.

5.1. SPACECRAFT ELEMENTS

Within the industrial consortium led by prime contractor Alenia Spazio, Astrium GmbH is responsible for the platform, Alcatel Space Industries (of France) are responsible for the EGG, and ONERA is responsible for the accelerometers and support to the performance assessment.

5.1.1. *Electrostatic Gravity Gradiometer (EGG)*

The objective of EGG is to measure the three components of the gravity-gradient tensor (*i.e.* gradiometer data). The EGG instrument which will be built at Alcatel incorporates accelerometers designed and developed at ONERA, and is based on an ambient temperature, closed loop, capacitive accelerometer concept. EGG is a three-axis gradiometer consisting of 3 pairs of three-axis servo-controlled capacitive accelerometers on an ultra-stable carbon-carbon structure. The thermal control (passive with heaters) provides 10 mK stability during 200 s. The performance is better than $3 \text{ mE Hz}^{-1/2}$ (see Table II). The EGG assembly has a mass of 150 kg and requires up to 75 W of electric power.

The principle of operation of the EGG is based on the measurement of the forces needed to maintain a proof mass at the centre of a cage. A six degree of freedom servo-controlled electrostatic suspension provides control of the proof mass in terms of translation and rotation. A pair of identical accelerometers, mounted on the ultra-stable structure 50 cm apart, form a “gradiometer arm”. The difference between accelerations measured by each of the two accelerometers, in the direction joining them, is the basic gradiometric datum (differential measurement), while half the sum is mainly proportional to the externally induced perturbing drag acceleration (common mode measurement). Three identical arms are mounted orthogonal to one another. The gradiometer axes so defined are nominally aligned in

TABLE II
Electrostatic Gravity Gradiometer (EGG) performance details

Measurement Band Width (MBW)	5×10^{-3} to 10^{-1} Hz
Baseline length	0.5 m
Accelerometer noise level in MBW	10^{-12} m s $^{-2}$ Hz $^{-1/2}$
Proof-mass positioning error	6×10^{-8} m Hz $^{-1/2}$
Absolute/relative scale factors	$10^{-3}/10^{-5}$
Absolute/relative misalignment	10^{-3} rad/ 10^{-5} rad

the along-track, cross-track and a third direction pointing approximately towards the Earth's centre (forming a right-handed triad). The three resulting differential accelerations provide direct, independent measurements: not only of the diagonal gravity components, but also of the perturbing angular accelerations (Rummel *et al.*, 2000).

In-orbit calibration of EGG involves carefully-planned, coordinated series of S/C manoeuvres and digital force-feedback information from the gradiometer to the Attitude and Orbit Control System (AOCS). Such calibrations will be repeated to check parameter stability with respect to thermal drifts and fluctuations. The objective of in-orbit calibration is to enhance the level of balancing to 10^{-5} in both scale-factor matching and alignment.

5.1.2. Satellite to Satellite Tracking Instrument (SSTI)

The objective of the SSTI is to provide support to the gravity field recovery, by using the positioning provided by the simultaneous tracking of up to 12 GPS satellite signals (in the SST-hl configuration). As such this payload element is an integral part of the system and not an independent instrument. In addition, the SSTI provides data for precise orbit determination and is used for real-time on-board navigation and attitude-reference-frame determination.

The selected Lagrange SST instrument has a redundant 12-channel dual-frequency receiver with a codeless tracking capability. It processes, demodulates and decodes the signals from GPS, received through a hemispherical antenna pointing in the zenith direction. The frequency bands L1 and L2 signals are used to allow the compensation of ionospheric delays by ground post-processing. Each channel of SSTI receives GPS signals and provides the following measurements: coarse acquisition pseudo range (L1; with provision for L2), L1 and L2 carrier phase (with phase noise < 1 mm), P1 and P2 code pseudo range (L1 and L2), L1-L2 differential carrier phase and P1-P2 differential pseudo range. In addition, the Lagrange SSTI provides the following capabilities:

- position and velocity measurements

- one pulse per second output synchronized with GPS time
- measurement time-tagging with respect to on-board spacecraft time
- fully redundant receiver and receiver processing unit
- optimisation of the number of measurement channels for power saving.

The total mass of the fully-redundant SSTI sub-system is approx. 15 kg, with a peak power demand of <40 W.

5.1.3. Laser Retro Reflector (LRR)

The LRR allows acquisition of a supplementary data set of satellite laser ranging (SLR) observations (by the existing SLR ground network) as backup for precise orbit determination post-processing. The LRR is a corner-cube array capable of reflecting laser pulses back along the incident light path.

5.1.4. Ion Thruster Assembly (ITA)

The ITA is Qinetiq's T5 Mk-5 dished-grid, magnetic field system. It consists of a quartz discharge chamber around which an RF field coil is wrapped, which induces the internal ionising electric field. Separate Xenon propellant flows feed the discharge chamber and a hollow-cathode neutralizer. A positive voltage on the screen grid attracts electrons into the discharge chamber from the neutralizer plasma, to initiate the discharge. A flat triple-grid system is used to extract the ion beam, with the thruster grid at +1200 V, the acceleration grid at -500 V, and a grounded deceleration grid. To minimize erosion, the acceleration grid is made from graphite. The ITA system on GOCE is operated in the drag control range; it goes from 100 W for 1 mN to 500 W for 12 mN. The 20 mN required for orbit reboost require 625 W of power input. The total mass of the ITA is approx. 60 kg.

The ITA system was already test-flown on the EURECA-1 mission (launch July 31, 1992–retrieval July 1, 1993). The ARTEMIS data relay satellite of ESA (launched in 2001) also employs an ITA propulsion system.

5.1.5. Standard Radiation Environment Monitor (SREM)

The objective is to provide radiation environment measurements. Since the EGG instrument of the primary payload is sensitive to electrical charging, SREM data can be used to correlate its measurements with encountered electron and proton fluxes. The ESA SREM instrument has already flown on the DERA mission STRV-1c (Space Technology Research Vehicle-1c). The SREM detector unit features two heads, each with a 20° half-cone field of view. The electronics unit comprises three particle detectors for electron and proton spectroscopy (measurement error <1%), cosmic-ray events counting and radiation-dose measurements. SREM has a mass of 2.6 kg and a power demand of 2.6 W.

5.2. SPACECRAFT DESIGN

The spacecraft consists of a long slender octagon structure, with a cross sectional area of approx. 0.9 m² and a length of approx. 5.0 m (Figure 2). It features total symmetry (about two planes) to minimize disturbances and there are no deployable appendages or moving parts. Within the structure there are several platforms upon which the payload modules are mounted, and which subdivide the platform into 3 modules for ease of integration. The lower module contains the AOCS/DFACS (Attitude and Orbit Control System/Drag-Free Attitude Control System) and Ion Thruster Assembly (ITA). The central module houses the EGG assembly and its electronics. In fact, the EGG assembly is located at the centre of mass of the spacecraft (and is required to stay within close proximity of the centre of mass throughout the spacecraft lifetime). The upper module largely contains the electrical equipment, data-handling and radio-frequency equipment, and houses the 40 kg Xenon propellant gas storage tank for the ion thruster.

Electric power is generated by four fixed body-mounted solar array panels and two wing-mounted solar panels (approx. 5.0 m²) with GaAs triple-junction cells. A Lithium Ion battery with 2.246 kWh total energy and 78 Ah capacity provide energy storage. The spacecraft thermal design and control is based on passive insulation and radiation techniques. The spacecraft has a launch mass of approx. 1000 kg, including propellant.

The key element of the AOCS/DFACS system on the GOCE spacecraft is the drag-free attitude control requirement. The DFACS is designed to compensate for the effects which atmospheric drag forces and torques have upon the gradiometer measurements using ion thrusters and micro-thrusters within the measurement bandwidth of 5 mHz to 100 mHz (see Canuto *et al.*, 2003). The total error budget for the gradiometer is of the order of 3 mE Hz^{-1/2} (Note: 1 E = 1 Eötvös = 10⁻⁹ s⁻², a unit of gravity gradient). In drag-free science operations mode, spacecraft attitude control (with an absolute pointing accuracy of 0.38 mrad) requires star tracker and EGG input data. However, attitude information is provided by an earth-sensor, sun-sensor and a three-axis magnetometer in other mission modes. Two 20 mN RF ion thrusters and eight pods of two micro-Newton proportional thrusters, based on field emission electric propulsion (FEEP), are used as actuators.

A conventional data handling system is used on-board, based on the MIL-1553 bus, an ERC 32 processor, and 2.5 Gbit data storage. Spacecraft communications are in S-band (two coherent S-band transponders, two antennas and a radio frequency distribution unit, 1 W RF power) with data rates of 2 kbit/s in the uplink and up to 850 kbit/s in the downlink. The ground receiving station is Kiruna and mission operations and control of GOCE will be conducted at ESOC.

The GOCE reference orbit is a sun-synchronous low-Earth orbit (altitude = 250 km; inclination = 96.5°), with a 06:00 hrs equatorial crossing, *i.e.* dawn-dusk orbit, or an equatorial crossing at 18:00 hrs, *i.e.* dusk-dawn orbit, at the ascending node (with the selection depending upon season of launch). Global coverage

outside the polar caps is reached after about 30–40 days. On this basis, a nominal mission duration of 20 months is planned for GOCE. A Rockot-class launch vehicle is used as the reference launcher.

6. GOCE Data Processing

The GOCE mission also requires a ground segment comprising reception and processing of the satellite telemetry data up to Level 1a/1b, managed by the Agency, and a higher-level data processing segment planned to be developed outside the Agency. Upon Delegations' request in the PB-EO, the Agency currently plans to include the Level 1 to Level 2 data processing within the scope of the GOCE Project activities, provided that the related additional funding is made available. A preparatory study has been conducted with the main objective of defining the architecture and the interfaces of the GOCE Level 1 to Level 2 processor. Based upon the outputs of this activity, and the available funding, the Agency will establish the framework within which it will manage the development of the Level 1 to Level 2 processor. This processor development is anticipated to take place in close cooperation with European centres of excellence with gravity field modelling and data processing expertise. Current plans exist to review aspects of ground segment design and the development of scientific data processing elements (level 1b to level 2) in Phase C/D.

The definitions for the different levels of data products are (ESA, 1996):

Level 0: time-ordered raw data as measured by GOCE. The satellite will downlink the data during contact with a dedicated ground receiving station.

Level 1a: instrument time series with the calibration data attached.

Level 1b: time series of calibrated and corrected instrument data along the orbit. These data include the primary instrument data: gravity gradients, SST-hl observations and GOCE satellite position; and other ancillary data such as the satellite linear and angular accelerations, satellite attitude, AOCS/DFAC thrust history, etc.

The two primary elements of the ground data processing chain are:

ESA Payload Data Segment (Level 0–Level 1a/b):

- mission performance assessment
- calibration and verification activities
- monitoring of the performance of the space segment and level 1b ground segment, feeding back to ESA Level 1a/b data products.

High Level Data Segment (Level 1–Level 2):

- processing to level 2 data products from level 1a /1b data
- precise orbit determination

- external calibration and validation.

Within the scope of the Level 2 data processor, the following three global products will be produced:

- gravity potential modelled as harmonic coefficients
- ground-referenced gridded values of geoid heights (Earth geoid map)
- global ground-referenced gridded values of gravity anomalies (Earth gravity map).

These Level 2 products will be accompanied by appropriate quality assessment products.

Level 2 data products are regarded as the starting point for further scientific analysis, and will be GOCE-only gravity-field solutions. Level 3 products are thus value-added, derivative or custom products developed for application in further studies of solid-earth physics, absolute ocean circulation, geodesy, sea-level rise etc. Additionally, products requiring combinations of surface or airborne gravimetric data, or other satellite or in-situ data with the GOCE data are regarded as Level 3 data products.

7. Conclusions

The GOCE mission will provide the first global, high spatial resolution and high accuracy observations of the Earth's gravity gradient tensor. These data will allow global and regional models of the (static) Earth's gravity field with unprecedented precision and spatial detail.

GOCE will obtain measurements with high spatial resolution and high (homogeneous) accuracy. Such a spatially detailed knowledge of Earth's gravity field will facilitate computation of an accurate equipotential reference surface, the "geoid" and quasi-exact orbit determination, particularly for low-earth orbiting satellites. The geoid as a reference surface is also directly applicable in applications such as levelling height determination using satellite techniques (known as 'GPS Levelling') and mapping of ocean and land surfaces.

New details of the Earth's gravity field provided by GOCE will also benefit a variety of Earth science disciplines. They will help to better understand processes that take place within the Earth's interior, and on and above its surface. Knowledge of the geoid, for instance, allows for studies of the solid Earth's mass distribution, interpretation of sea-level changes, ocean water flows/ocean heat transport and related with these, climate studies and model predictions.

In summary, GOCE data will undoubtedly find broad application in the fields of geodesy, oceanography and solid-earth physics.

Acknowledgements

Significant contributions of the GOCE Phase A Study participants and GOCE Mission Advisory Group, together with those of the GOCE Project and Industrial consortium members; Alenia Spazio, Astrium GmbH, Alcatel Espace, and ONERA are gratefully acknowledged. Sincere thanks go to Rudolf von Steiger, the organisers of the Workshop on *Earth Gravity Field from Space* and all its participants for making it such a milestone event in this decade of gravity field missions.

References

- Canuto, E., Martella, P., and Sechi, G.: 2003, 'Attitude and drag control: an application for the GOCE satellite', *Space Sci. Rev.*, this volume.
- European Space Agency (ESA): 1986, 'SESAME: Solid Earth Science & Application Mission for Europe', ESA Special Workshop Proceedings, ESA SP-1080, 128pp.
- European Space Agency (ESA): 1991, 'The Solid-Earth Mission ARISTOTELES', International Workshop Proceedings, ESA SP-329, 137pp.
- European Space Agency (ESA): 1996, 'Gravity Field and Steady-State Ocean Circulation Mission', ESA SP-1196(1), Report for Assessment of the Nine Candidate Earth Explorer Missions, 77pp.
- European Space Agency (ESA): 1998, 'The Science and Research Elements of ESA's Living Planet Programme', ESA SP-1227, 105pp.
- European Space Agency (ESA): 1999, 'Gravity Field and Steady-State Ocean Circulation Mission', Report for Mission Selection, ESA SP-1233(1), 217pp.
- European Space Agency (ESA): 2001, <http://www.esa.int/livingplanet/goce>.
- Rebhan H., Aguirre, M., Johannessen, J. A.: 2000, 'The Gravity Field and Steady-State Ocean Circulation Explorer Mission — GOCE', *Earth Observation Quarterly* **66**, July 2000, 6–11.
- Rummel, R.: 1979, 'Determination of short-wavelength components of the gravity field from satellite-to-satellite tracking or satellite gradiometry; an attempt to an identification of problem areas', *Manuscripta Geodetica* **4**, 107–148.
- Rummel, R., Müller, J., Oberndorfer, J. H., and Sneeuw, N.: 2000, in R. Rummel, H. Drewes, W. Bosch, and H. Hornik (eds.), *Satellite Gravity Gradiometry with GOCE, Towards an Integrated Global Geodetic Observing System (IGGOS)*, IAG Section II Symposium, 1998, Springer Verlag, pp. 66–72.
- Wolff, M.: 1969, 'Direct Measurement of the Earth's Gravitational Potential Using a Satellite Pair', *J. Geophys. Res.* **74**, 5295–5300.

EARTH GRAVITY FIELD FROM SPACE – FROM SENSORS TO EARTH SCIENCES: CLOSING REMARKS

G. BEUTLER

Astronomical Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

Received: 17 January 2003; Accepted in final form: 22 January 2003

1. Introduction

In his treatise “the Theodicy” Gottfried Wilhelm Freiherr von Leibniz (1646–1716) wrote in 1710 that *...I have not omitted to show that it is a consequence of the supreme perfection of the Sovereign of the Universe, that the kingdom of God be the most perfect of all possible states or governments, and that consequently the little evil there is, is required for the consummation of the immense good which is there found ...* (Leibniz, 1710). Usually this quote is today used in the more condensed form saying that *we are living in the best possible of all universes*.

Statements of comparable weight concerning the current status of geodesists and geodesy were made by Reiner Rummel in his opening address of the symposium *Earth Gravity Field from Space – From Sensors to Earth Sciences*: after more than thirty years of continuous struggles for the deployment of one dedicated gravity mission, three of them, namely CHAMP (CHAllenging Minisatellite Payload), GRACE (Gravity Recovery And Climate Experiment), and GOCE (Gravity field and steady-state Ocean Circulation Explorer) are in orbit (CHAMP and GRACE) or about to be deployed (GOCE). The workshop participants may even remember that Reiner Rummel used the term “geodesist’s paradise” for the upcoming “decade of geopotentials” (Sneeuw, 2002). The written version (Rummel, 2002) of the opening address is slightly less explicit (as it is often observed in such cases).

The latter remark also applies to the written version of these closing remarks: the thoughts developed subsequently are only loosely related to what was actually said on the last day of the workshop. We believe, however, that it is important to check in retrospect, whether the following central issues were addressed (explicitly or implicitly) at the workshop:

1. Are we actually entering a new era in geosciences with the three dedicated space missions?
2. What are the key results expected of the new era?
3. Are there conclusions and/or predictions for a next phase (or even a next era) in this field?



4. Is there “little evil” (Leibniz, 1710) in the best possible of all geodetic worlds, which seemingly is about to begin?

2. The “Old” Era of Gravity Field Mapping

When advocating the advent of a “new” era one should at least mention how the preceding era looked like. In our case the preceding era is that of determining the Earth’s gravity field from ground-based observations of the *geocentric orbits* of artificial Earth satellites, which are described by the equations of motion

$$\mathbf{R} \dot{\mathbf{r}} + V \cdot \mathbf{r}/C :: ; \quad (1)$$

where \mathbf{r} is the geocentric radius vector of the satellite, $V \cdot \mathbf{r}$ /the Earth’s gravitational potential, and $\dot{\mathbf{r}}$ the gradient operator; the ellipsis symbolizes all the forces (per unit mass) acting on the satellite, which are *not* related to the gravitational attraction exerted by the Earth. Atmospheric drag, solar and albedo radiation pressure, but also the gravitational attraction by Sun and Moon are contributing to this part.

In order to remind ourselves of how the geopotential $V \cdot \mathbf{r}$ /was actually determined in this earlier era we recall that symbolically it may be written as

$$V \cdot \mathbf{r} / D G \int_{V_e}^Z \frac{\% \cdot \mathbf{r}_{dV_e} /}{\mathbf{r} \cdot \mathbf{r}_{dV_e}} dV_e ; \quad (2)$$

where G is the constant of gravitation, V_e the volume occupied by the Earth (solid part, oceans, and atmosphere), dV_e one of the Earth’s volume elements, and $\% \cdot \mathbf{r}_{dV_e} /$ the density of matter within dV_e .

According to potential theory the Earth’s potential may be developed into a series of harmonic functions

$$V \cdot \mathbf{r} ; , ; / D \frac{GM}{r} \sum_{i=0}^{\infty} \frac{a_e}{r} \sum_{k=0}^{\infty} P_i^k \cdot \sin \gamma / f C_{ik} \cos k, C S_{ik} \sin k, g ; \quad (3)$$

where a_e is the equatorial radius of the Earth, r , γ , and γ are the spherical geocentric coordinates of the satellite’s geocentric position vector; the functions $P_i^k x$ / are the associated Legendre functions, and the C_{ik} and S_{ik} are coefficients of the development to be determined. When speaking of gravity field determination we mean the determination of these coefficients.

Usually the equations of motion are set up in a quasi-inertial geocentric system (a system which is always parallel to an inertial system, but which cannot be inertial because it is geocentric). Equation (1) would then imply that the development (3) also refers to the same quasi-inertial system. Technically, one might proceed in this way. The drawback resides in a very pronounced time dependence of the coefficients C_{ik} and S_{ik} . It makes therefore more sense to refer the development (3)

to a system co-rotating with the Earth, to evaluate the gradient of the potential in the same system, and to transform the resulting force vector back into the inertial system. The Earth orientation and rotation parameters are needed as functions of time for that purpose (for the time interval of interest).

The coefficients of the geopotential might be determined “easily”, if the geocentric orbit $\mathbf{r}(t)$ could be accurately reconstructed from the observations without making use of the equations of motion (1). In the sense of a thought experiment one could easily establish the trajectory in this way: One merely needs a very high density of sites observing the trajectory continuously without errors! Provided the trajectory is available, one might now set up and solve a non-linear parameter estimation problem with the initial state vector $\mathbf{r}(t_0)$; $\mathbf{P}(t_0)$ and the C_{ik} and S_{ik} as unknowns using the coordinates of the position vectors $\mathbf{r}(t)$ as pseudo-observations. The practical solution differs considerably from this simplified view, but this does not matter for our considerations.

In reality the satellites’ trajectory could not be established well in this earlier era of gravity field determination: Optical observations (astrometric places) were used initially (from the early 1960s to the mid 1970s) for this purpose, LASER observations were almost uniquely used after 1975 for the same purpose. The latter technique usually is referred to as SLR, Satellite Laser Ranging. The SLR observations are globally coordinated by the International Laser Ranging Service (ILRS, <http://ilrs.gsfc.nasa.gov>). The key problem with this approach is the sparsity of observations: Due to the limited number of stations in the global ILRS network and due to weather conditions it is not possible to achieve (close to) continuous observations of the satellites’ orbits.

Despite all these deficiencies, this “old” era of gravity field determination was a true revolution in itself, when it started in the early 1960s: Whereas only the Earth’s oblateness, i.e., the term C_{20} of the expansion (3), was known with reasonable accuracy in the pre-satellite era, hundreds of terms could be firmly determined in this phase of satellite geodesy. By combining terrestrial with satellite geodetic observations it was possible to establish developments of type (3) which were truncated only after the terms of degree and order of about $i \leq k \dots 70$.

Satellites such as the two LAGEOS satellites (standing for LASer GEodetic Satellites) were paramount for the exploration of the Earth’s stationary gravitational field using the SLR technique. Even the temporal (secular) change of a few of the terms of the Earth’s potential could be determined. The determination of the secular drift C_{20} and an unexpected anomaly thereof near the end of the twentieth century, which could be extracted from more than thirty years of LAGEOS observations (Cox and Chao, 2002), is just one example of the achievements of that epoch.

3. Innovations Emerging From CHAMP, GRACE, and GOCE

Gravity field determination with CHAMP (CHAllenging Minisatellite Payload) is conventional in the sense that the coefficients are determined solely from solving the equations of motion (1). The innovation is due to

- | the continuous tracking of CHAMP's orbit using the Global Positioning System (GPS) and
- | the use of accelerometers to determine independently the non-gravitational forces acting on the satellite.

The first aspect virtually removes the key difficulty associated with the conventional determination of the gravity field by SLR and other optical methods, namely that of data sparseness. Also, thanks to the very low altitude of CHAMP, its orbit is much more sensitive to the subtleties of the Earth's gravity field than the orbits of conventional LASER satellites.

Using different approaches it is (or will be) possible to reconstruct the geocentric trajectory of Low Earth Orbiters (LEOs) like CHAMP, GRACE, and GOCE on the (few) cm level. Hugentobler (2002) and Visser (2002) discuss the reconstruction of the trajectory using the GPS in their contributions. This approach is, by the way, very close to the result of our thought experiment outlined in Sect. 2.

The second aspect is essential when using LEOs for gravity field determination. It is absolutely mandatory to separate the non-gravitational forces from the gravitational ones in the case of rather bulky LEO space vehicles. In all the new missions this separation is achieved through accelerometers inside the spacecraft by measuring in essence the displacement of a test mass, which is shielded in the satellite's interior against the prominent surface forces acting on the satellite. Accounting for the non-gravitational forces in the term “...” in Eq. (1) using the accelerometer measurements basically allows us to determine the gravity field parameters without introducing biases. Alternatively, the satellite motion may be compensated in real time for the non-gravitational accelerations.

The CHAMP results presented by Reigber et al. (2002) are truly impressive: The exploitation of *only three months* of CHAMP data drastically improves the so-called long-wavelength gravity field recovery (compared to the results achieved in about forty years of conventional geopotential determinations). Results of this type truly justify the label “new era of gravity field mapping”.

It is, however, also interesting to note that a “blind” application of the CHAMP-derived gravity field to other (than CHAMP) orbits is not trivial. According to Reigber et al. (2002) a combination with observations related to these other orbits is still required. This had to be expected: Any particular orbit is specifically sensitive to specific terms of the geopotential, namely the resonance terms. Also, small systematical errors of the accelerometers are inevitably mapped into small biases of the gravity field parameters – and such biases can only be detected by using the determined geopotential for modelling independent orbits.

GRACE and GOCE add one more essential aspect – the experts probably would say *the* essential aspect – to the new era of gravity field recovery, namely that of gradiometry. Whereas CHAMP (and all previous methods) reconstruct the geopotential by solving the equations of motion (1) for one or more satellites, GRACE and GOCE continuously measure in addition one (or more) second derivative(s) of the Earth's potential. As these in situ measurements may be established with an extremely high precision and rate, the contribution of gradiometry to the determination of the geopotential usually dominates the contribution due to the reconstruction of the satellites' orbits (this is in particular true for the higher frequency part of the forces' spectra).

In the case of GRACE the gradiometric method is implemented by observing and analyzing the difference between the orbits of two co-orbiting satellites (GRACE A and B), in the case of GOCE by a set of six accelerometers inside the spacecraft.

The two GRACE satellites may be viewed as a highly accurate one-dimensional gradiometer oriented (more or less) in along-track direction, whereas the six accelerometers in GOCE form a complete three-dimensional gradiometer, which provides all six independent second-order derivatives of the geopotential. As both, the GRACE and the GOCE satellites are orbiting in near-polar orbits, the gradiometers scan virtually the entire Earth, allowing us to reconstruct the potential function (3) almost perfectly. The accelerometers onboard GRACE and GOCE also allow us to separate gravitational from non-gravitational effects.

Modelling the geopotential in this new era of gravity field determination is a demanding task of applied mathematics. Sneeuw (2002) gave a concise overview of the techniques which are under development for analyzing the data of the three space missions. The peculiarities associated with the GRACE gradiometer are discussed by Balmino (2002).

4. The Workshop in Retrospective

In Sect. 3 we argued that we are actually entering a new era in gravity field mapping. Our arguments were technical in nature. The technical aspects of the new era of gravity field mapping were actually dealt with in the first session of the workshop. Before resuming the questions asked initially, it may be appropriate to recall briefly the structure of the workshop “Earth Gravity Field from Space – From Sensors to Earth Sciences”. It consisted of the following six sessions:

1. Precise Orbit Determination and Gravity Field Modelling
2. Solid Earth Physics
3. Ocean Circulation
4. Geodesy
5. Sea Level
6. Future Concepts

The first technical session was followed by the four application sessions, illustrating how the four fields of Earth sciences benefit from the new era of gravity field mapping. It does not make sense to list here all the titles of the presentations (this may be taken from the table of contents of this volume), nor does it make sense to review the individual contributions in detail. We will mention below a few examples to support the perception that the new era of gravity field determination also may be viewed as a new era for (at least some of) the Earth sciences – in particular when dealing with global aspects.

The last session of the workshop was devoted to the discussion of future concepts. We will review some material from this session in the concluding section.

5. A New Era in Earth Sciences

The explanation of the acronyms for the GOCE and the GRACE mission, namely

- | Gravity field and steady-state Ocean Circulation Explorer (GOCE) and
- | Gravity Recovery And Climate Experiment (GRACE)

indicate that both missions are supposed to have a significant impact on branches of science other than geodesy. It is even safe to state that the two missions never could have been realized as “geodesy-only” missions.

The three space missions CHAMP, GRACE and GOCE truly will revolutionize our knowledge of and understanding for the Earth’s outer gravity field, the Earth’s interior, the Earth’s atmosphere and oceans, and the development in time of the entire system. The latter aspect is sometimes described by the well-established slogan of *global change*. This was the clear outcome of the application sessions 2–5. Why is this the case? Let us use two illustrations.

The simple, very general defining equation (2) for the Earth’s gravitational potential $V \cdot \mathbf{r}$ /acting on a satellite with the geocentric position vector \mathbf{r} , is helpful to answer this question: Whereas it is *not* possible to reconstruct the Earth’s density distribution unequivocally by determining only its potential (2) (and/or its first two derivatives) outside the mass distribution, it *may*, on the other hand, be possible to determine the continuous mass redistribution within the system, *provided* we have additional information. This is the case when studying the planet’s water cycle: Water is, e.g., evaporated from the sea surface, it is transported via clouds to the continents, where it resides for some time as a snow cover or as ground water, and eventually flows back to the oceans. This redistribution of water is (in essence) constrained to the Earth’s surface, which allows us to interpret the measured changes of the geopotential rather easily in terms of a (surface) mass redistribution.

The global water cycle is one of the key justifications for the GRACE mission and explains (in part) the second half “CE” of the GRACE experiment. As a complete set of gravity field parameters will be determined with every monthly batch of GRACE data, it should be easily possible to extract quasi-periodic signals of the order of one year. Wahr and Velicogna (2002) even argue that the GRACE

mission, provided it is operational for at least five years, should allow it to study post-glacial rebound phenomena. There can be no doubt that secular trends in a significant number of geopotential terms will be firmly established by GRACE. It will be in addition interesting to see whether anomalies of the kind reported by Cox and Chao (2002) are likely to be real or rather artifacts.

Whereas GRACE aims at determining temporal variations in the gravity field and to interpret them (e.g., as the planet's water cycle), GOCE aims at the determination of the best possible stationary gravity field (naturally including the determination of secular changes of many terms). Once this field is determined, e.g., in the form (2), it is a relatively straight forward affair to construct equipotential surfaces. The geoid, the equipotential surface at sea level, is of particular interest in this context. (The reader is by the way invited to consult Vermeer (2002) or Sacerdote and Sansò (2002) to see how "straight forward" such tasks are). The geoid is the reference surface to which the sea surface topography, as obtained from satellite altimeter missions, has to be referred. The height of the oceans above or below the geoid is the essential quantity from which ocean circulation models (and possible changes thereof) can be deduced. This aspect, the combination of sea surface topography and gravity field maps, was probably the principal driver for the GOCE mission and it is represented by three letters (Ocean Circulation Experiment) in the GOCE acronym. Woodworth and Gregory (2002) show what kind of insight concerning the sea level problem is expected from GOCE (and GRACE). The expected contribution to the solution of the sea level problem by the GRACE mission was addressed by Nerem et al. (2002).

The two examples (water cycle and stationary gravity field) show that the two missions GOCE and GRACE are fundamental for all future studies related to the sea level and to the Earth's water cycle. Results of the kind expected by the workshop participants truly justify the advocation of a new era in some of the Earth sciences, which is made possible by the sequence of the gravity-related space missions CHAMP, GRACE and GOCE.

6. Conclusions

The material and arguments presented so far more or less support the short version of the quote taken from Leibniz (1710): Geodesy and Earth sciences really enter a new, very promising era thanks to the gravity-related space missions. Let us focus here on the last of the questions asked initially: Is there "little evil" in the best possible of all geodetic worlds?

First of all we have to slightly correct the perception that the three space missions could/can be realized without additional knowledge. Nothing could be more wrong! CHAMP, GRACE, and GOCE absolutely need the best possible

- | realization of the celestial reference system,
- | realization of the terrestrial reference system,

- | orbits and satellite clock corrections of the GNSS (Global Navigation Satellite Systems) used to precisely position the LEOs, and
- | the best possible transformation parameters between the celestial and the terrestrial systems during the lifetime of the missions.

The best realization of the celestial reference system is the ICRS, the International Celestial Reference Frame, which is defined and maintained by the International Earth Rotation Service (IERS) in collaboration with the IVS, the International VLBI (Very Long Baseline Interferometry) Service for Astrometry and Geodesy. The best realization of the terrestrial reference system is the ITRF, the International Terrestrial Reference Frame, which is set up and maintained by the IERS in collaboration with the ILRS, the IVS and the IGS (International GPS Service). The ITRF does not only consist of a list of station coordinates, but also of the “velocity” field associated with it. The orbits and clocks of the GNSS satellites are made available through the IGS and its specialized analysis centers. The accuracy is on the few centimeter level for the satellite coordinates, on the few picosecond level for the satellite clocks. Currently “only” GPS satellites are used. The Galileo system (ESA system) and the GLONASS (Russian system) will be used as soon as orbits and satellite clocks of a reasonable accuracy will become available.

It is therefore fair to state that the new era of satellite gravity field mapping is based on an existing and well organized infrastructure set up by the IAG, the International Association of Geodesy.

Another important issue, if not the central issue, resides in the separation of gravitational and non-gravitational forces. We simply assumed so far that this separation is elegantly achieved by the spaceborne accelerometry. Accelerometers are, however, instruments with a (hopefully well known and well defined) error spectrum. Interestingly enough it is the low-frequency part that is critical. This implies in turn that the low-order terms will be mainly determined by the satellite orbit (and not only by gradiometry). In order to make the geopotential determination via orbit reconstruction (the CHAMP case) accurate, one needs on the other hand a (close to) perfect separation of gravitational and non-gravitational forces. For this purpose one needs again the accelerometers – which were, however, already said to have comparatively poor error characteristics for low frequencies. This may (in part) explain the insufficiency of the CHAMP-only gravity field to model the orbits of the LASER satellites (see Sect. 3).

In view of these considerations it is not surprising that alternative designs for the currently used electrostatic accelerometers were proposed in session 6, e.g., by Zarembiński (2002) and Aguirre-Martinez and Sneeuw (2002) to improve the GOCE-type gradiometers. Bender and Klipstein (2002), on the other hand, wish to improve the performance of between-satellite links by using optical techniques, and Bender et al. (2002) propose to even further improve the GRACE-concept by replacing the one-dimensional gradiometer (realized by two spacecrafts) by multi-dimensional gradiometers (to be realized by a constellation of spacecrafts).

In his opening address (Rummel, 2002) proposes that strategies for future missions should be based on the optimization of three variables, namely the experiment height, the compensation of the attenuation of the gravity signal by differential measurements, and the measurement precision.

The session on future concepts did not bring forward completely different principles (comparable to the advent of gradiometry in the new era). *Cum grano salis* the improvements may be characterized by the famous Olympic trinity *altius, citius, fortius*.

It will certainly take the scientific community some time to properly process and interpret the data produced by the three missions. It is moreover safe to predict that unexpected features will be observed in the gravity field (and its evolution in time) determined from the sequence of new mission. It will be extremely difficult or even impossible to separate secular signals from a long periodic signal (long compared to the mission duration). The author of these lines would therefore not be amazed to see the experts in the field advocate in the near future the necessity for a (practically) continuous monitoring of the Earth's gravity field by the state-of-the-art gradiometry and satellite positioning tools available at the times of the mission. Such developments are rather common in geodesy, in particular in recent years. The scientific services of IAG (monitoring and maintaining the reference frames and the corresponding transformation parameters) but also altimetry missions are excellent examples. The CHAMP mission (when compared to the old era of gravity field mapping) also proves that the transition from a sporadically available to a continuous observation scenario may improve results by orders of magnitude.

It seems that we are about to enter the “best possible of all geodetic worlds”, but that we will have a lot of work to minimize “the little evil there is” – but according to Leibniz this is even a necessity.

Acknowledgements

The workshop participants wish to express their deep gratitude towards the International Space Science Institute (ISSI) for hosting the Workshop *Earth Gravity Field from Space – From Sensors to Earth Sciences*. The ISSI facilities, the professionalism of the team, and the warm hospitality experienced by the workshop participants were paramount for the success of the workshop. A special tribute goes to Dr. Ruedi von Steiger, who was the spirit throughout the workshop and the editing and reviewing process. Only a broad-band scientist with remarkable administrative talents is capable of assuming this key role in such a modest, yet effective manner.

References

- Aguirre-Martinez, M. and N. Sneeuw: 2002, 'Needs and tools for future gravity measuring missions'. This Volume.
- Balmino, G.: 2002, 'Gravity field recovery from GRACE: Unique aspects of the high precision inter-satellite data and analysis methods'. This Volume.
- Bender, P., J. L. Hall, J. Ye, and W.M. Klipstein: 2002, 'Satellite-satellite LASER links for future gravity missions'. This Volume.
- Bender, P. J., R.S. Nerem and J.M. Wahr: 2002, 'Possible future use of LASER gravity gradiometers'. This Volume.
- Cox, C.M. and B.F. Chao: 2002, 'Detection of a large-scale mass redistribution in the terrestrial system since 1998'. *Science*, Vol. 297, 831.
- Hugentobler, U. and G. Beutler: 2002, 'Strategies for precise orbit determination of low Earth orbiters using the GPS'. This Volume.
- Leibniz, G.W.: 1710, 'Essais de théodicée sur la bonté de dieu, la liberté de l'homme et l' origine du mal'. *Amsterdam: chez Isaac Troyel*.
- Nerem, S., J.M. Wahr and E.W. Leuliette: 2002, 'Measuring the distribution of ocean mass using GRACE'. This Volume.
- Rummel, R.: 2002, 'How to climb the gravity wall'. This Volume.
- Reigber, Ch., G. Balmino,, P. Schwintzer, R. Biancale, A. Bode, J.-M. Lemoine, R. König, S. Loyer, H. Neumayer, J.-Ch. Marty, F. Barthelmes, F. Péoranz, and S. Y. Zhu: 2002, 'Global gravity field recovery using solely GPS tracking and accelerometer data from CHAMP'. This Volume.
- Schuh, W.-D.: 2002, 'The processing of band-limited measurements: filtering techniques in the least squares context and in the presence of data gaps'. This Volume.
- Sacerdote, F. and F. Sansò: 2002, 'Remarks on the role of the height datum in altimetry-gravimetry boundary value problems'. This Volume.
- Sneeuw, N.: 2002, 'Space-wise, time-wise, torus and Rosborough representations in gravity field modelling'. This Volume.
- Vermeer, M.: 2002, 'The elusive stationary geoid'. This Volume.
- Visser, P.: 2002, 'Aiming at the 1-cm orbit for low earth orbiters: reduced dynamic and kinematic precise orbit determination'. This Volume.
- Wahr, J. and I. Velicogna: 2002, 'What might GRACE contribute to +studies of post glacial rebound'. This Volume.
- Woodworth, P.L. and J.M. Gregory: 2002, 'Benefit of GRACE and GOCE to sea level studies'. This Volume.
- Zarembiński, S.: 2002, 'On superconductive gravity gradiometry in space'. This Volume.

Address for Offprints: G. Beutler, Astronomical Institute, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland, gerhard.beutler@aiub.unibe.ch

Author Index

- AGUIRRE-MARTINEZ, M., **407**
 AMBE, D., 193
- BAKER, T., 205
 BALMINO, G., **7**, 55
 BARTHELMES, F., 55
 BEGGAN, C., 205
 BENDER, P.L., **377**, **385**
 BEUTLER, G., 17, **433**
 BIANCALE, R., 55
 BIELLA, G., 115
 BINGHAM, R.J., 205
 BINGLEY, R., 205
 BODE, A., 55
 BOUMAN, J., **293**
 BREMOND, M., 163
- CABANES, C., 131
 CANUTO, E., **357**
 CAZENAVE, A., **131**
- DAVIDSON, F., 239
 DAVIS, J.L., 81
 DE FRANCO, R., 115
 DOMINH, K., 131
 DRINKWATER, M.R., **419**
- EGBERT, G.D., 271
- FLOBERGHAGEN, R., 419
- GENNERO, M.C., 131
 GREGORY, J.M., 307
- HAAGMANS, R., 419
 HAINES, K., **205**
 HALL, J.L., 377
 HERNANDEZ, F., 205, 239
 HIPKIN, R., 205
 HOLT, J., 205
 HUGENTOBLER, U., **17**
- HUGHES, C.W., **217**
- ICHIKAWA, K., 195
 IMAWAKI, S., **195**
- KÖNIG, R., 55
 KLIPSTEIN, W.M., 377
 KNUDSEN, P., **261**
 KOOP, R., 293
- LE PROVOST, C., 131, **163**
 LE TRAON, P.Y., **239**
 LE GRAND, P., **225**
 LEMOINE, J.-M., 55
 LEULIETTE, E.W., 331
 LOYER, S., 55
- MAROTTA, A.M., **95**
 MARTELLA, P., 357
 MARTY, J.-C., 55
 MILNE, G.A., 81
 MITROVICA, J.X., 81
 MUZI, D., 419
- NEREM, R.S., **331**, **385**
- NEUMAYER, H., 55
- PEROSANZ, F., 55
 POPESCU, A., 419
- RAY, R.D., **271**
 REIGBER, C., **55**
 RIO, M.H., 239
 ROWLANDS, D.D., 271
 RUMMEL, R., **1**
- SACERDOTE, F., **253**
 SANSÒ, F., 253
 SCHRAMA, E.J.O., **179**
 SCHUH, W.-D., **67**

- SCHWINTZER, P., 55
SECHI, G., 357
SNEEUW, N., 37, 409
STAMMER, D., 147
STEPANOV, V., 217
SWENSON, S., 345
- TAMISIEA, M.E., 81
TONDI, R., 115
TOUBOUL, P., 393
- UCHIDA, H., 195
- VAN DEN IJSSEL, J., 27
VELICOGNA, I., 319
VERMEER, M., 283
VERMEERSEN, L.L.A., 105
VISSER, P.N.A.M., 27
- WAHR, J.M., 319, 331, 345, 385
WOODWORTH, P., 307
WUNSCH, C., 147
- YE, J., 377
- ZAREMBIŃSKI, S., 367
ZHU, S.Y., 55

List of Participants

- M. Aguirre-Martinez, *ESA ESTEC*, Miguel.Aguirre@esa.int
 G. Balmino, *CNES, GRGS, Toulouse*, Georges.Balmino@cnes.fr
 P.L. Bender, *University of Colorado at Boulder*, pbender@jila.colorado.edu
 G. Beutler, *University of Bern*, beutler@aiub.unibe.ch
 R. Biancale, *CNES, GRGS, Toulouse*, Richard.Biancale@cnes.fr
 A. Cazenave, *Observatoire Midi Pyrénées*, anny.cazenave@cnes.fr
 B.F. Chao, *NASA GSFC*, chao@bowie.gsfc.nasa.gov
 K. Danzmann, *University of Hannover*, danzmann@mpq.mpg.de
 R. de Franco, *CNR-DPA, Milano*, defranco@irrs.mi.cnr.it
 J.O. Dickey, *NASA JPL / Caltech*, jean.o.dickey@jpl.nasa.gov
 M.R. Drinkwater, *EOP-SMO, ESA/ESTEC, Noordwijk*, Mark.Drinkwater@esa.int
 R. Floberghagen, *EOP-PGM, ESA/ESTEC*, Rune.Floberghagen@esa.int
 R. Haagmans, *EOP-SME, ESA/ESTEC*, Roger.Haagmans@esa.int
 K. Haines, *Reading University*, kh@mail.nerc-essc.ac.uk
 U. Hugentobler, *University of Bern*, urs.hugentobler@aiub.unibe.ch
 C.W. Hughes, *Bidston Observatory*, cwh@pol.ac.uk
 S. Imawaki, *Kyushu University*, imawaki@riam.kyushu-u.ac.jp
 P. Knudsen, *Kort & Matrikelstyrelsen*, pk@kms.dk
 R. Koop, *SRON National Institute for Space Research*, R.Koop@sron.nl
 C. Le Provost, *Observatoire Midi-Pyrénées*, christian.le-provost@cnes.fr
 P.Y. Le Traon, *CLS, Pierre-Yves.Letraon@cls.fr*
 P. Le Grand, *IFREMER*, plegg@ifremer.fr
 A.M. Marotta, *University of Milano*, anna.maria.marotta@unimi.it
 P. Martella, *ALENIA Spazio, Torino*, pmartell@to.alespazio.it
 A. Milani, *University of Pisa*, milani@dm.unipi.it
 J. Müller, *University of Hannover*, mueller@ife.uni-hannover.de
 S. Nerem, *University of Colorado*, nerem@ccar.colorado.edu
 R.D. Ray, *NASA GSFC*, richard.ray@gsfc.nasa.gov
 R. Rummel, *TU München*, rummel@bv.tum.de
 R. Sabadini, *Università degli Studi di Milano*, roberto.sabadini@unimi.it
 F. Sansò, *Politecnico di Milano*, fsanso@ipmtf4.topo.polimi.it
 H. Schmeling, *J.W. Goethe Universität Frankfurt a.M.*, Schmeling@geophysik.uni-frankfurt.de
 E.J.O. Schrama, *Delft University of Technology*, schrama@geo.tudelft.nl
 W.D. Schuh, *University of Bonn*, schuh@theor.geod.uni-bonn.de
 P. Schwintzer, *Geo-Forschungsztzentrum Potsdam*, psch@gfz-potsdam.de
 N. Sneeuw, *University of Calgary*, sneeuw@geomatics.ucalgary.ca
 W. Spakman, *Utrecht University*, wims@geo.uu.nl
 D.B. Stammer, *University of California at San Diego*, dstammer@ucsd.edu
 M.E. Tamisiea, *University of Toronto*, tamisiea@physics.utoronto.ca
 V. Toumazou, *CNES/Legos-GRGS*, Vincent.Toumazou@cnes.fr
 M. Vermeer, *Helsinki University of Technology*, martin.vermeer@hut.fi

- B. Vermeersen, *Delft University of Technology*, b.vermeersen@lr.tudelft.nl
P. Visser, *Delft University of Technology*, Pieter.Visser@lr.tudelft.nl
J. Wahr, *University of Colorado*, wahr@lemond.colorado.edu
P. Woodworth, *Bidston Observatory*, plw@unixa.nerc-bidston.ac.uk
C. Wunsch, *MIT*, cwunsch@pond.mit.edu
S. Zarembiński, *Chalmers University of Technology*, f4asz@fy.chalmers.se

Space Science Series of ISSI

1. R. von Steiger, R. Lallement and M.A. Lee (eds.): *The Heliosphere in the Local Interstellar Medium*. 1996 ISBN 0-7923-4320-4
2. B. Hultqvist and M. Øieroset (eds.): *Transport Across the Boundaries of the Magnetosphere*. 1997 ISBN 0-7923-4788-9
3. L.A. Fisk, J.R. Jokipii, G.M. Simnett, R. von Steiger and K.-P. Wenzel (eds.): *Cosmic Rays in the Heliosphere*. 1998 ISBN 0-7923-5069-3
4. N. Prantzos, M. Tosi and R. von Steiger (eds.): *Primordial Nuclei and Their Galactic Evolution*. 1998 ISBN 0-7923-5114-2
5. C. Fröhlich, M.C.E. Huber, S.K. Solanki and R. von Steiger (eds.): *Solar Composition and its Evolution – From Core to Corona*. 1998 ISBN 0-7923-5496-6
6. B. Hultqvist, M. Øieroset, Goetz Paschmann and R. Treumann (eds.): *Magnetospheric Plasma Sources and Losses*. 1999 ISBN 0-7923-5846-5
7. A. Balogh, J.T. Gosling, J.R. Jokipii, R. Kallenbach and H. Kunow (eds.): *Co-rotating Interaction Regions*. 1999 ISBN 0-7923-6080-X
8. K. Altwegg, P. Ehrenfreund, J. Geiss and W. Huebner (eds.): *Composition and Origin of Cometary Materials*. 1999 ISBN 0-7923-6154-7
9. W. Benz, R. Kallenbach and G.W. Lugmair (eds.): *From Dust to Terrestrial Planets*. 2000 ISBN 0-7923-6467-8
10. J.W. Bieber, E. Eroshenko, P. Evenson, E.O. Flückiger and R. Kallenbach (eds.): *Cosmic Rays and Earth*. 2000 ISBN 0-7923-6712-X
11. E. Friis-Christensen, C. Fröhlich, J.D. Haigh, M. Schüssler and R. von Steiger (eds.): *Solar Variability and Climate*. 2000 ISBN 0-7923-6741-3
12. R. Kallenbach, J. Geiss and W.K. Hartmann (eds.): *Chronology and Evolution of Mars*. 2001 ISBN -7923-7051-1
13. R. Diehl, E. Parizot, R. Kallenbach and R. von Steiger (eds.): *The Astrophysics of Galactic Cosmic Rays*. 2001 ISBN 0-7923-7051-1
14. Ph. Jetzer, K. Pretzl and R. von Steiger (eds.): *Matter in the Universe*. 2001 ISBN 1-4020-0666-7
15. G. Paschmann, S. Haaland and R. Treumann (eds.): *Auroral Plasma Physics*. 2002 ISBN 1-4020-0963-1
16. R. Kallenbach, T. Encrenaz, J. Geiss, K. Mauersberger, T.C. Owen and F. Robert (eds.): *Solar System History from Isotopic Signatures of Volatile Elements*. 2003 ISBN 1-4020-1177-6
17. G. Beutler, M.R. Drinkwater, R. Rummel and R. Von Steiger (eds.): *Earth Gravity Field from Space – from Sensors to Earth Sciences*. 2003 ISBN 1-4020-1408-2