Lars E. Sjöberg
Mohammad Bagherbandi

Gravity Inversion and Integration

Theory and Applications in Geodesy and Geophysics

Springer

Gravity Inversion and Integration

Lars E. Sjöberg · Mohammad Bagherbandi

Gravity Inversion and Integration

Theory and Applications in Geodesy and Geophysics

Springer

Lars E. Sjöberg
Division of Geodesy and Satellite Positioning
Royal Institute of Technology
Stockholm
Sweden

Mohammad Bagherbandi
Division of Geodesy and Satellite Positioning
Royal Institute of Technology
Stockholm
Sweden

and

Department of Industrial Development, IT and Land Management
University of Gävle
Gävle
Sweden

ISBN 978-3-319-50297-7 ISBN 978-3-319-50298-4 (eBook)
DOI 10.1007/978-3-319-50298-4

Library of Congress Control Number: 2016963159

Cover figure: The cover figure shows a global map of the geoid undulation based on the EGM2008 gravity model and the GRS80 reference ellipsoid (resolution 1×1 arc-degree). The figure was created by the authors. For more details see Fig. 1.5.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This textbook provides a comprehensive overview of gravity integration and inversion, which contributes to physical geodesy and geophysics, and it identifies classical and modern topics for studying the Earth. It discusses both theoretical and practical aspects, e.g., for the determination of a precise geoid model besides presenting ample worked examples. Physical geodesy terminology is used throughout the book. The unprecedented knowledge of the Earth's gravity field and its temporal variation are progressively capturing the attention of the geosciences for many reasons. As a result of recent dedicated satellite missions, knowledge of the global to regional gravity field has reached extraordinary levels of quality and resolution. The modeling of the Earth's mass distributions in the crust and its interior, as well as the temporal changes/transports of such masses, is most important in studying geodynamics. The enhanced knowledge of the 3D-layered structure of the Earth will improve our capability to understand, monitor and predict geophysical processes, which potentially threaten our technically developed society. Today, thanks to the development of atomic clocks, the original idea of A. Bjerhammar from 1975 of chronometric leveling for direct measurement of geopotential differences is emerging as a fascinating new tool in geodetic and geophysical applications.

The aim of this book is to provide students at the M.Sc and Ph.D. levels, as well as researchers, basic and some in-depth knowledge about the current and recent theory and application of gravity for geodesy and geophysics, as seen mainly from a geodesist's perspective and with an emphasize on theory. Physical geodesy is treated rather generally, and a main goal is to provide the reader with a theory (the KTH method) for determining "the 1-cm geoid" (including both geoid and quasigeoid methods), while the geophysical applications of gravity are limited to the determination of crustal depth and density contrast at the crust/mantle boundary, and stress and viscosity in the upper mantle, as well as some simple examples of how to estimate the mass and depth of some other large-mass structures in the Earth from combinations of the geopotential, gravity, and gravity gradients. Temporal changes of the gravity field are treated with emphasis on long-term trends, while periodic changes are more sparsely utilized.

Another motivation behind this book is, to honour the centennial anniversary of, and, to some extent, announce and revive some research ideas originating with the deceased A. Bjerhammar (1917–2011) at the Royal Institute of Technology (KTH) in Stockholm, Sweden. Bjerhammar is well-known for his research in physical geodesy and geodetic studies on the Fennoscandia land-uplift phenomenon and its relation to the regional gravity field and upper mantle viscosity in the region, as well as for being a pioneer in developing a theory for linear adjustment of erroneous observations using generalized matrix algebra. Unfortunately, much of his research can only be found in disparate papers presented in various journals, at conferences and as internal reports from KTH. As the first author of this book has been a M.Sc and Ph.D. student of Bjerhammart from 1969 to 1975, his co-worker until 1980, and his successor since 1984, this book to some extent sought to reflect and further continue his research ideas, as well as on other ideas developed and partly published in papers, etc.

The second author was a Ph.D. student at KTH during 2007–2011 and a postdoc 2011–2014, with the prime author as his supervisor on the determination of the crustal depth and density contrast from Earth gravitational models using an isostatic model as a specialty in his research. He currently performs multidisciplinary research and is active in several directions, such as in geophysics, geodesy and land surveying (applied geodesy). His main research interest is to develop and study relations between geodesy and geophysics especially study on temporal changes of the Earth's gravity field and glacial isostatic adjustment.

Stockholm, Sweden — Lars E. Sjöberg
March 2017 — Mohammad Bagherbandi

Acknowledgements

We acknowledge that Professor Artu Ellmann and Associate Professor Jonas Ågren spent part of their summer vacation in 2016 to read a draft version of this book. Their valuable comments significantly improved the final manuscript.

We thank the following publishers and journals for granting permissions to reuse some previously published material of our own: Springer for articles in *Acta Geodaetica et Geophysics Hungary, Pure and Applied Geophysics, Surveys in Geophysics, Encyclopedia of Geodesy* and *Journal of Geodesy, Studia Geodaetica et Geophysica, Surveys in Geophysics* and also de Gruyter for articles in *Journal of Geodetic Science* and *Journal of Applied Geodesy*, and, finally, Elsevier for articles in *Earth Science Review* and *Physics of the Earth and Planetary Interiors.*

Thanks also go to Petra van Steenbergen, Executive Editor in Earth Sciences, Geography and Environment, and Hermine Vloemans, project coordinator, at Springer for their help to prepare the book.

Finally we acknowledge the Division of Geodesy and Satellite Positioning at Royal Institute of Technology (KTH) for providing a good working environment and the opportunity to complete this book.

Contents

Chapter 1
Introduction

Abstract The task of this book is to convey some basic and in-depth knowledge on the traditional and recent theory of gravity inversion and integration for applications in geodesy and geophysics. The subject fields, physical geodesy and geophysics including geodynamics with emphasis on gravity information, are briefly introduced. Also the recent history of physical geodesy and gravimetric geophysics is reviewed. It follows that satellite techniques are indispensable for global to regional studies. The basic but important concepts of a geodetic reference system, frame and datum are also defined with examples. The Geodetic Reference System 1980, a versatile tool in most applications of gravity data, is presented in some detail.

Keywords Geoid · Reference system · Reference frame · Geodynamics

1.1 Contents of the Book

Chapter 2 provides the basic mathematics for this book. Here least squares theory, coordinate systems, spherical and ellipsoidal harmonics, fundamental potential theory and regularization of improperly posed problems and/or numerically singular systems of equations are introduced. Chapter 3 reviews classic physical geodesy, including its integral formulas, height systems and their determinations. Chapter 4 presents modern physical geodesy starting with the original concepts of M.S. Molodensky. A major part of this chapter is a variety of modifying Stokes' formula for geoid computation by combing terrestrial gravity data and an Earth Gravitational Model (EGM). Chapter 5 deals with the topographic, atmospheric and ellipsoidal corrections needed in Stokes' formula. Chapter 6 expresses and compares today's commonly used methods for modifying Stokes' formulas for geoid and quasigeoid determination: the Remove-Compute-Restore (RCR) and Least Squares Modification of Stokes' formula with Additive corrections (LSMSA) techniques. A number of applied geoid projects with these techniques are also reported. In Chap. 7 the descriptions of several modern tools in physical geodesy are gathered. These include quasigeoid determination, corrections from quasigeoid to geoid, combined

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_1

geoid determination from gravimetric and geometric data, the determination of the potential of the geoid, geoid determination by spectral combination of satellite, airborne and terrestrial gravity data, and, finally, it introduces a relativistic approach to determine geopotential and height differences originating with Bjerhammar (1975). Chapter 8 comprises methods for gravity inversion as well as analyses for temporal changes of the gravity field.

1.2 The Subject Field

Gravity is a fundamental universal force. At the Earth's surface, gravity (g) is on average 9.81 m/s^2, and it reduces with distance squared from the centre of the Earth when moving into space. Due to the flattening of the Earth towards the poles, gravity there is 0.5% greater than at the equator, and mass density variations inside the planet and its topography cause additional regional variations. Gravity also changes with time due to mass transport inside the Earth. Large, slow secular mass motions caused by phenomena such as mantle convection and glacial isostatic adjustment yield gravity changes on the order of a few parts in 10^{-9} of g, while the amplitudes of seasonal variations caused by the hydrological cycle on Earth are on the same order of magnitude. Ocean tidal effects give rise to several periods with greatest amplitudes for the daily tide that can reach more than 16 m in the Bay of Fundy, Canada, while the body tide is considerably smaller. All these gravity variations can be observed locally by accurate absolute and relative gravimeters, and in particular by superconducting gravimeters, while recent satellite gravity missions, such as the twin satellites of Gravity Recovery And Climate Experiment (GRACE) and its follow-on mission, planned to be launched in 2017, enable regional and global studies.

Physical geodesy and geophysics share the endeavour of using the Earth's gravity field as an indispensable observable in various geoscience studies and for a variety of technical enterprises, e.g. in land and engineering surveying and construction, hydrography, geophysical prospecting and mining. While the physical geodesist is primarily interested in determining the Earth's size and shape and mapping the exterior gravity field, the geoid and the deflections of the vertical, the geophysicist uses the gravity field to explore, explain and interpret various geophysical and geologic processes and internal mass structures of the Earth. Most of such analyses require *integration* of gravity over the whole globe or at least over a region. Today such analyses are frequently facilitated by the use of global EGMs, usually determined by satellite data or from a combination with terrestrial gravity data. The employment of satellite data at the Earth's surface or in the Earth's interior implies some type of *gravity inversion*, and *inverse problems* are closely related with improperly posed problems. Gravity inversion, i.e. to determine some features/structures in the Earth's interior by gravity, is an *ambiguous* operation, as the solution for any such structure is not unique without additional information for its solution. Already the downward continuation of satellite gravity data to the

Earth's surface is an inverse problem. More so when performing harmonic continuation of the gravity field from external space to the Earth's interior, which inevitably introduces a bias, as the gravity field is not harmonic inside the topographic masses, and this problem should be treated properly. In this way, the task to determine the geoid, at least what concerns the continental geoid, shares with most gravity related problems in geophysics being an inverse problem.

Among the variety of tools for exploring the Earth's interior, gravity data is comparatively cheap and easy to collect, traditionally by terrestrial observations but today also regionally and globally by utilizing airborne and satellite gravity missions.

While local applications are usually best treated in the spatial domain and the integration and/or inversion processes are typically performed in flat earth approximations, global studies gain from the long-wavelength EGMs expressed in spherical harmonics, which leads to computations in the spectral domain. Frequently the derivations of suitable integration and inversion formulas are conveniently utilized in the frequency domain, while the resulting formulas may or may not be applied in the space domain as integral formulas or integral equations or as combinations of spectral series and integral formulas. It is obvious that the data provided by an EGM today yields long- to rather short-wavelength information about the gravity field, while local details can only be achieved by dense terrestrial gravity and gravity related data. The question how to combine such diverse data sets for optimal solutions is essential, and various strategies are at hand. This is particularly important in the case of utilizing emerging ultrahigh degree EGMs.

Since the advent of satellite geodesy in 1957, global models of the Earth's gravity field of reasonable quality could finally be determined from satellite orbit tracking and by using advanced satellite sensors and inhomogeneously distributed geodetic observatories around the world. The problem with this strategy for resolving the global gravity field is that the satellite missions were not designed for gravity field determination, leading to non-optimal solutions. Therefore, for the first time, dedicated satellite gravity missions, launched about the last millennium shift, made a giant step forward in determining the global gravity field accurately and homogeneously in the low to medium wavelength bands. Then the question arises whether such data, collected repeatedly for a longer period of time, can also be used for studying the gravity changes with time. Starting with satellite laser ranging over about 20 years, which records of data successfully detected the static and secular change of the very long-wavelength gravity field, today the dedicated satellite gravity mission GRACE has proved that space geodesy is capable of determining not only the static gravity field from space to higher degrees of accuracy and homogeneity, but, for the first time, temporal changes to higher degrees are reliably determined down to a few degrees of resolution, enabling the estimation of various causes of mass changes within the Earth system, of great importance in understanding Earth processes and climate change.

The Earth's gravity field is a basic geophysical parameter that reflects the Earth's mass distribution and its changes with time. As the Earth is not a rigid body but is

constantly changing under the influence of external and internal forces (such as mantle convection, plate tectonics, sea level rise, etc.), the gravity field and its changes with time are of great significance for studying various geodynamic and geophysical processes.

The Earth's temporal deformations can be divided into secular, periodic and episodic (suddenly accelerating or decelerating) phenomena. Our knowledge about the causes and effects of these deformations vary, and, as one example, frequently a long periodic deformation is interpreted as a secular one. Some phenomena such as Earth tide and post-glacial rebound have been observed systematically for long time and are relatively well understood. Figure 1.1 illustrates major time variable gravity field sources in terms of spatial resolution, and almost all of these sources can be significantly sensed by the GRACE data. One can see the time and spatial extensions and resolutions of each phenomenon, but it should be emphasized that, because of limitations in the resolution of observations, not all but the majority of them can be studied by the present GRACE data. As stated above, the temporal changes of gravity are caused by secular, periodic and episodic phenomena, but frequently the causative sources belong to two or all three types of phenomena. For example, mantle convection, plate tectonics and Glacial Isostatic Adjustment (GIA) are typically secular causes, while Earth and ocean tides appear both as static gravitational effects and as

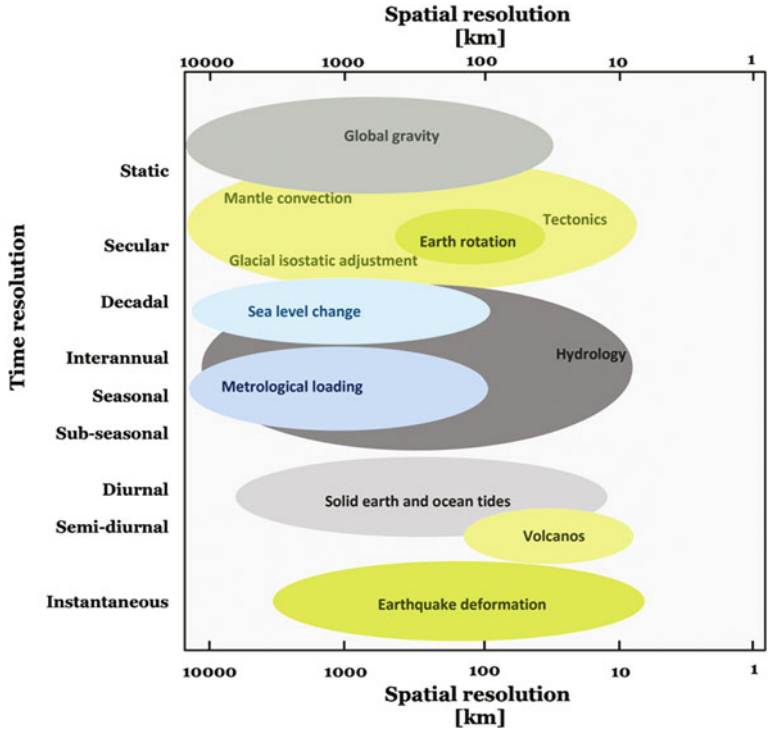

Fig. 1.1 Time variable gravity field sources in terms of spatial resolution (Moritz 1980a)

periodic signals at periods from decades to seconds, and earthquakes (usually followed by regional tectonic deformations and tsunamis) are episodic. Therefore, the geodetic measuring techniques relevant to studying such changes will differ. For instance, to understand mantle convection and plate tectonics, the global gravity field is a suitable observable, while for studying problems related to GIA gravity, also vertical crustal motions and sea level changes are suggested sets of data.

Generally the Earth responds to deforming forces as a visco-elastic medium, implying that the response is dependent on the temporal scale and spatial extension of the deformation. See Fig. 1.1. For short periodic deformations, the Earth behaves as an elastic medium, while for secular changes its responses are visco-elastic to viscous.

1.3 The Development of the Subject Field Before the Last Millennium Shift

Geodesy as practiced by the ancient Greeks (geometric geodesy) belongs, together with astronomy, to the oldest of the known sciences. In contrast to geometric geodesy, physical geodesy and gravimetric geophysics are young sciences with roots in Galileo Galilei's (1564–1642) and Isaac Newton's (1642–1727) fundamental discoveries in the gravity field, primarily learned through their experiments with pendulums. Newton also predicted the geometric flattening of the Earth, which was confirmed in the 1790s based on the results of the French grade measurement expeditions to Lapland and Peru. Fortunately, the phenomenon of isostasy (see below), which was not discovered until the end of the Great Trigonometrical Survey in India (from 1802 to 1871), did not significantly affect the result. Only in the late 1800s, when the gravimeter had been developed to useful accuracy, both in theory and in the field, applied research were carried out. Nevertheless, theoretical developments started early, and the most important formula in physical geodesy, Stokes' formula, was published by George Gabriel Stokes in 1849. However, as geoid determination by Stokes' formula requires that gravity data be integrated, in principle, all over the Earth, the formula could not be used in practice for a long time. Similar to most other applications of gravity in physical geodesy, extensive gravity networks are needed, which is not possible to accomplish until general gravity surveys have been conducted.

Similarly, many geophysical problems, such as those related to GIA and plate tectonics, are regional to global in nature, and consequently require gravity data from vast regions for their solutions. The deeper inside the Earth the geophysical phenomenon under study has its origin, the more that long-wavelength gravity data is essential.

The last 60 years have seen an enormous development in the assessment of gravity and related data on the Earth's surface and in space. Since the advent of the first artificial satellite Sputnik in 1957, it has finally been possible to accurately determine the long- to medium-wavelength parts of the gravity field to high accuracy, allowing successful global geodetic and geophysical studies. Kaula

(1966) published a very timely textbook on gravimetric satellite geodesy, which theory is still in use for determining the long- to medium-wavelengths of the gravity field in spherical harmonics. Kaula also predicted the power spectrum of gravity on Earth ("Kaula's rule"; see Eq. 3.29), which is still valid. All types of available ground based geodetic satellite methods were used in studying the perturbations of satellite orbits, in particular for recovering "the Stokes coefficients" in the form of EGMs. These techniques suffered from a number of errors due to the heterogeneity of observation coverage over the globe and the low sensitivity of higher degree harmonics, yielding unstable systems of equations for solving not only the lowest degrees. However, by conditioning the solutions by Kaula's rule and using a combination of terrestrial data, the upper degree of the EGMs could be advanced to higher degrees, while the quality of the significant low- to medium-degrees were still relatively moderate. A general treatise on the theory of physical geodesy was Heiskanen and Moritz (1967), which was updated by Hofmann-Wellenhof and Moritz (2006).

During this period, M.S. Molodensky's new and revolutionary theory in physical geodesy became known to geodesists in the West (Molodensky et al. 1962). In particular, the concepts of the surface gravity anomaly, normal height and height anomaly/quasigeoid, enabling the determination of the shape of the Earth without knowing the density distribution of the topography, were most remarkable. Bjerhammar (1962) presented his version for determining the quasigeoid from a finite number of surface gravity anomalies by analytical continuation to an internal sphere ("the Bjerhammar sphere"), but this was still not a well recognized method (and many geodesists are still reluctant to do so). At that time, the only influential geodesists who fully appreciated this technique were H. Moritz (Heiskanen and Moritz 1967, Sects. 8–10) and M. Hotine (Hotine 1969, p. 318). Somewhat later, Molodensky's new concept of modifying Stokes' formula (see Sect. 4.4) was also recognized.

In preparing his report on the fundamental theory of physical geodesy, Krarup (1969) set out to prove that Bjerhammar's method could not be applied as a reliable approximation technique. However, during the course of his work, he became fascinated with the method, developed it further to the method of least squares collocation (LSC) and, finally, came up with a theoretical proof of the validity of analytical continuation in physical geodesy (the Runge-Krarup theorem). Finally, Moritz (1980a) published his textbook *Advanced Physical Geodesy*, partly covering the topic of least squares collocation and partly methods for accurate quasigeoid determination.

Bjerhammar et al. (1980) estimated the viscosity of the (upper) mantle in Fennoscandia using a spectral window of the Earth's gravity field. Kaula (1972) suggested a completely different approach to study the relation between gravity and viscosity in the mantle, namely, if a heavy slab sinks into the mantle (which is likely the case in subduction zones) and the viscosity is constant, then the slab should drag light crustal material down along with it, yielding negative gravity anomalies. This technique was confirmed by Hager (1984) and Richards and Hager (1984). In contrast they also found that, if the viscosity has a sharp increase at mid-mantle depth (say, 670 km), the gravity signal is weakly positive.

Turcotte and Schubert (2002) introduced the reader to the basics of geophysics and geodynamics, such as plate tectonics, stress and strain in solids, elasticity and flexure, heat transfer, fluid mechanics, rock rheology faultings, etc. The chapter on gravity is of special interest, as it introduces some methods for gravity integration and inversion with many numerical examples. For treatments of a variety of theories and geophysical applications of gravimetry, the following volumes could be mentioned. Lambeck (1988) deals with various applications of geodetic tools for geophysics/geodynamics with in-depth descriptions, while Moritz (1990) provides a theoretical/mathematical treatment of the Earth's interior, equilibrium figures, the equipotential ellipsoid and its density distribution and isostasy. Vaníček and Christou Nikolaos (1994) present several areas of application of gravity for geophysics. Watts (2001) provides an in-depth presentation of isostasy and flexure of the lithosphere, both from theoretical points of view and from geological and geophysical findings.

1.4 Recent Developments in Gravimetric Theory and Data

1.4.1 Development of Gravimetric Data

Although terrestrial gravity data have been gradually measured for covering and densifying land areas, the event of satellite altimetry in the late 1970s was the most important advance in terrestrial data coverage, as it enables gravity mapping of the oceans, which was lacking before. Today, satellite altimetry covers all oceans with detailed gravity related information, which are invaluable data sets as gravity was very poorly known at sea in the past. Also, repeated, time tagged altimetry tracks provide valuable information on temporal changes of sea level. In addition, during the last decade, large regions on land and at sea have been covered by airborne gravity data. All these new measurements are complemented with much more accurate positioning by Global Navigation Satellite Systems (GNSS). In particular, the vertical position, in the past frequently poorly measured, is now accurately determined by GNSS.

In parallel, the long-wavelength gravity field has been explored to unprecedented resolution and accuracy, first by dynamic satellite geodesy methods using and combining all kinds of geodetic-satellite measurements, and finally by the break through that came between 2000 and 2015 with dedicated satellite gravity missions. Already during the 1980s and 90s, there were competing technologies for such missions, such as NASA's Gravity Research Mission, based on satellite-to-satellite tracking technique (e.g. Wolff 1969; Sjöberg 1982; Jekeli and Upadhyay 1990) and ESA's ARISTOTELES mission using satellite gradiometry, but it was not until the millennium shift that such gravity measurement tools were realized (through CHAMP, GRACE and GOCE satellite missions). In this way, the long-wavelength gravity field has been covered to a resolution of, say, 100×100 km^2, providing gravity and geoid accuracies of the orders of 1 mGal and 1 cm, respectively.

In the meantime, satellite data were also combined with terrestrial gravity data to create high resolution EGMs. For example, EGM2008 (Pavlis et al. 2012), complete to degree and order 2160, corresponds to a resolution of the order of 5×5 km^2. In parallel, also high resolution Digital Elevation Models have been developed based on recently acquired terrestrial data, as well as space data. Examples of such data bases are DTM2006 (Pavlis et al. 2007) and ETOPO1 (Amante and Eakins 2009) to resolutions of 5×5 and 1×1 km^2, respectively.

Since the introduction of GPS for civilian uses in 1983, the world has seen a formidable development of GNSS technology and methods, which, on one hand, provides an independent geometric technique for validating gravimetric geoid models, and, on the other hand, calls for more accurate geoid models as the reference level for GNSS and for directly providing geoid height estimates by the so-called GNSS-levelling technique. Thirdly, GNSS is an essential tool in dedicated satellite gravity missions.

1.4.2 Development of Theory

Advances in scientific results cannot only rely on improved data sets, but theory must also improve. As discussed in the previous section, up until the 1980s, the goal for the accuracy in geoid determination was of the order of several decimetres, and the emphases in available literature (like Moritz 1980a; Sansó and Sideris 2013) were and are on quasigeoid rather than geoid determination. However, as the development of the quality and amount of gravimetric data advanced, geodesists more and more strive for the "1-cm geoid", and the need for more accurate theory becomes obvious. This was emphasized by Martinec (1998), who developed several improved practical formulas for accurate geoid modelling according to traditional Stokes-Helmert approach. There are also numerous scientific papers that develop physical geodetic theory in various ways, while recent textbooks on the subject matter are missing. In this context should be mentioned that the development of the Earth Gravitational Model EGM2008 (Pavlis et al. 2012) is a comprehensive achievement that combines theory and data in such a way that Earth gravitational models are taken to a new level of accuracy and resolution. Nevertheless, more efforts are needed to reach the 1-cm goal.

During the last decades, there has been a revolution in the Earth sciences. The study of the Earth's interior (crust, mantle and core) and attempts to model it in 3-D is still a major challenge in geophysics, in particular for modelling the physical and chemical structures. Moreover, mapping and understanding geodynamic processes like deformations of the Earth's surface and its interior mass structures, often related to global change, are developing. In most of these studies, satellite methods significantly simplify and enhance regional to global compilations and results. Although the progress in geophysics is rich and manifold, this study is focused on the following gravity related topics.

The last 15–20 years has seen many global and regional compilations of the Earth's crustal depth, mainly based on seismic data. As an alternative, this study will present recent mathematical methods based on isostasy, which utilize satellite-gravity data and isostatic methods for determining global and regional crustal depth models, as well as models of the Moho density contrast. Refined models are obtained by combinations of seismic and gravimetric data.

Already in the 1960s, Runcorn (1967) suggested a mathematical technique to determine tectonic stress in the mantle from gravity. Until recently, numerical applications of the method were not very successful. However, based on recent high quality satellite gravity data, we will demonstrate that this method has its merits.

For a long time, the Fennoscandia land uplift phenomenon as part of the global GIA has been a rich field for geodynamic studies, mainly because the area is easily accessed and has the long records of geological data needed for modelling the historical ice sheet development and sea level progress with time. Through the years, a number of global GIA models were developed and tested on land uplift records in Fennoscandia (e.g. Steffen and Wu 2011). Most important parameters in these tests are the lithosphere thickness and the viscosities in the upper and medium mantle. Consequently, viscosity profiles in the mantle have lately been a major research field by means of such models. Major uncertainties in this type of modelling are related to the ice thickness and extension history, sea level rise and parameters of the Earth's rheology. In this study, the static gravity field and its temporal change are employed as an alternative technique to determine the upper–mantle viscosity.

Textbooks like Fowler (2001) and Turcotte and Schubert (2002) introduce the reader to the basic concepts in geophysics and, in particular, geodynamics by describing major Earth processes with many mathematical examples. The dedicated satellite gravity missions deliver data that open the possibility to determine secular and periodic variations of the long to medium wavelengths of the gravity field, which enable the realization of a number of geodynamic studies. In the present book, we will demonstrate how to apply the information on temporal changes of gravity to determine the secular and annual variation of the geoid. The change of the geoid in Fennoscandia, related to GIA, will also be used to estimate the upper to medium mantle viscosity.

Recently Jacoby and Smilde (2009) provided numerous examples of the geological and geophysical interpretations of gravity data. They suggest that gravity interpretation be related to broader subjects such as the shape of the Earth, the nature of the continental and oceanic crust, isostasy, forces and stresses, geological structure, climate change, etc. For example, one of the important applications of observing gravity in geodesy is to use it in combination with other measurements, e.g. distances and coordinates for defining the Earth's shape and in combination with levelling and other surveying methods (GPS levelling). In geophysics and geology, the purpose is typically for the exploration of the Earth's interior, and gravity data has a similarly important bearing on oceanography. In geodynamics, temporal gravity change has become an interesting matter because its space time

behaviour reflects processes such as loading or unloading and flow inside the Earth, which depend on viscosity and mantle convection. Temporal gravity variation can also reveal mechanical properties and even deep processes such as Earth core oscillations (Jacoby and Smilde 2009).

The dedicated satellite gravity missions play and will play increasing roles in studying and explaining the Earth's interior constitution and its on-going geodynamic processes, and, in the next few years, one can expect to see fairly detailed 3-D maps of the Earth based on the inversion and integration of gravity, seismic and other geophysical data. Also, temporal changes of sea level and gravity are key issues in studying changes in ocean circulation, which are closely related with climate change.

1.5 Reference System, Reference Frame and Datum

Geospatial data becomes a more and more important tool in society for many kinds of research of immediate use, but also for future planning and enterprise. For the latter purpose, geospatial data archives deserve increasing attention. In this respect, the data archives will form virtual models of the Earth, and the modelled Earth will be the basis, e.g. for future engineers, researchers and decision makers for their applications, predictions and decisions. From such a perspective, one needs to clearly define and distinguish between the basic geodetic concepts of *reference system*, *frame* and *datum*. It is important to clarify the difference between reference system and reference frame, as they are two different concepts. The former should be regarded as a theoretical definition, and the latter is the practical counterpart realized through observations and a set of coordinates.

A *reference system* defines models, parameters and constants, which serve as necessary bases for the mathematical representations of the involved geometric and physical quantities. A geodetic *reference system* is the joint concept of a coordinate system, a time system, gravity model and a number of physical constants and datum parameters. Reference systems used in satellite geodesy are global, 3-D and geocentric, i.e. with origin of the coordinate system at the gravity centre of the Earth (various terrestrial systems, which could be 3-D, 2-D or 1-D, with various origos). Two types of systems are needed/relevant in satellite geodesy: Earth-fixed and space-fixed systems.

A geodetic *reference frame* is a realization of a reference system, i.e. it is the resulting practical system based on methods of observation and analyses of the data, as well as the orientation of the coordinate system. Practically, this is done *physically* by a solid materialization of points, and, *mathematically*, by determining the parameters, e.g. geometric coordinates (Drewes 2009). Hence, the reference frame makes possible the determination of station location/position, possibly as a function of time. This procedure seems simple, but one has to deal with complicated matters such as theory of relativity, forces acting on the satellites, corrections for the atmosphere, Earth rotation, solid Earth and ocean tides, tectonic motions, etc.

The geodetic datum explains clearly the relation between a reference system and a reference frame by assigning some parameters, e.g. the coordinates of the origin of the system (X_0, Y_0, Z_0), the directions of coordinate axes and the scale as a unit of length (e.g. in metre).

The definition of a reference system must not be affected by the realization of a geodetic datum and reference frame. Also, the errors of measurements in a reference frame do not affect the datum definition, as the datum parameters are fixed and independent of the measurements. The roles of the reference system and frame in the sustainable development of a country have been reaffirmed by the United Nations General Assembly Resolution on the Global Geodetic Reference Frame, which states that:

> The economic and scientific importance of and the growing demand for an accurate and stable global geodetic reference frame for the Earth that allows the interrelationship of measurements taken anywhere on the Earth and in space, combining geometric positioning and gravity field related observations, as the basis and reference in location and height for geospatial information, which is used in many Earth science and societal applications, including sea level and climate change monitoring, natural hazard and disaster management and a whole series of industrial applications (including mining, agriculture, transport, navigation and construction) in which precise positioning introduces efficiencies (United Nations 2015).

The resolution further invites Member States to:

> Commit to improving and maintaining appropriate national geodetic infrastructure as an essential means to enhance the global geodetic reference frame, engage in multilateral cooperation that addresses infrastructure gaps and duplications towards the development of a more sustainable global geodetic reference frame and develop outreach programmes that make the global geodetic reference frame more visible and understandable to society (United Nations 2015).

Geodetic reference systems and frames can be global, regional or local. It may be appropriate to use a national reference system when small details over a large area should be recognized and a more locally adapted reference system for large scale information.

Examples of Geodetic Reference Systems

A well-defined spatial reference system is needed to create an accurate digital geospatial data base or map. The reference system defines the coordinate system and datum in which all landmarks have unique (3-D) coordinates. As an example, the International Terrestrial Reference System (ITRS) constitutes a set of prescriptions and conventions together with the modelling required to define origin, scale, orientation and time evolution of a Conventional Terrestrial Reference System (CTRS). For more details, see, e.g. http://www.ggos-portal.org (retrieved 13 March 2016).

There is a mutual dependence existing between national and international reference systems. National systems need to have a close connection to the international systems, which needs observations from the national reference system. Therefore, international cooperation has intensified over the past decade, and,

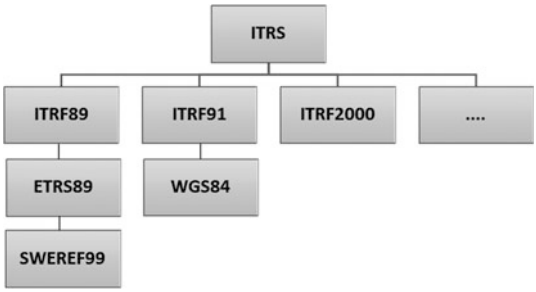

Fig. 1.2 Relations between the national and international reference systems ITRS, ITRF, ETRS and SWEREF99

among other things, has led to the reference systems becoming realizations equivalent to European systems. For example, within Europe, ETRS89 (European Terrestrial Reference System) has been adopted as the joint reference system. ETRS89 is compatible with the global ITRS, which is a dynamic system that includes a velocity model. Coordinates and speeds in ITRS are calculated based on data from various time intervals, and these solutions are referred to as ITRF (International Terrestrial Reference Frame). Figure 1.2 shows the relationship between global systems and SWEREF99 (the national reference system of Sweden).

Examples of Geodetic Reference Frames

The ITRS is realized by the International Terrestrial Reference Frame (ITRF) based upon estimated coordinates and velocities of a set of stations observed by the geodetic space techniques VLBI, LLR, GPS, SLR and DORIS. A fixed geodetic reference frame is a necessary tool to better understand time dependent phenomena, and it is also needed for precise orbit determination of geodetic satellites. A terrestrial reference frame provides a network of reference points located on the Earth's surface with known coordinates. The Earth's motion in space and the deformation and motion of points on its surface are measured vs. the International Celestial Reference Frame fixed by VLBI, LLR, GPS, SLR and DORIS techniques. Figure 1.3 shows the temporal changes in the ITRF2008 station positions (Altamimi et al. 2011).

"Geodetic reference frames are the basis for three-dimensional, time dependent positioning in global, regional and national geodetic networks, for spatial applications such as the cadastre, engineering construction, precise navigation, geoinformation acquisition, geodynamics, sea level and other geoscientific studies."

The geodetic reference frames are necessary to consistently reference or tag parameters using geodetic observations, e.g. station coordinates, crustal motion, Earth orientation information, etc. Ground observations of GNSS and other satellites, or radio telescope observations of distant quasars, enable us to define not only the reference frame, but also to derive other parameters, for instance, crustal motion parameters from continuously operating GNSS stations, Earth orientation parameters, and geocentric motion (for more, see http://www.ggos-portal.org).

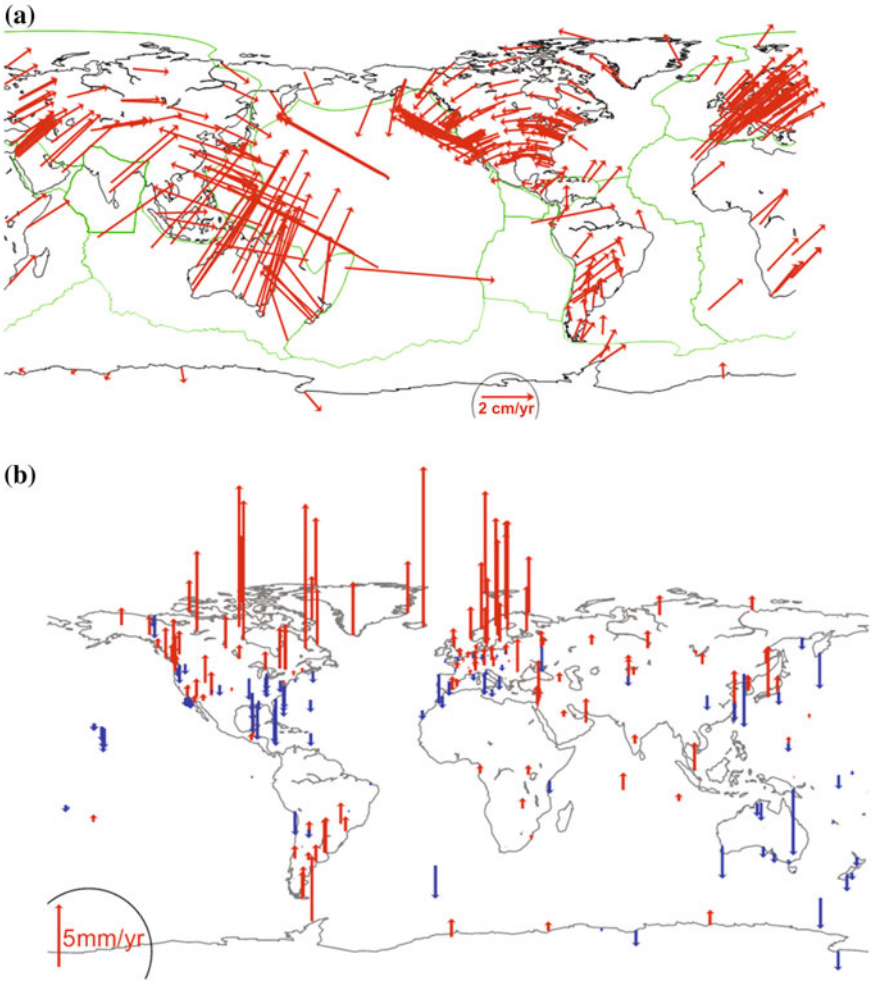

Fig. 1.3 International terrestrial reference frame (ITRF) 2008 realization stations and its **a** horizontal (Major plate boundaries are shown according to Bird 2003) and **b** vertical velocity field. Positive velocities are shown in *red* and negative in *blue* (Altamimi et al. 2011) with formal error less than 0.2 mm/yr

"The Terrestrial Reference frame provides a set of coordinates of some points located on the Earth's surface. It can be used to measure plate tectonics, regional subsidence or loading and/or used to represent the Earth when measuring its rotation in space. This rotation is measured with respect to a frame tied to stellar objects, called a celestial reference frame."

The ITRF is being updated from time to time, and, for example, the notation ITRF2014, the year implies a realization with computation of the frame at epoch 2014. Each realization consists of station-position coordinates and their velocities.

1.5.1 More on Reference Systems

The location of a point can be expressed in several different coordinate systems within the same reference. As an example, a point in GRS80 can be specified with geocentric coordinates or geodetic coordinates in any of the projection zones.

In the geodetic description of the Earth's surface, there are three different surfaces; see Fig. 1.4:

- The Earth's physical surface, which also includes sea level.
- The geoid, which is the equipotential surface of the Earth's gravity field that best fits to the mean sea level. (An equipotential surface is a surface of constant scalar potential. The gradient of the gravity potential is called gravity, and it is perpendicular to the surface.)
- A reference ellipsoid, which is a mathematical description of the ellipsoid that best fits the geoid in a region. The globally best fitting ellipsoid is denoted the Mean Earth Ellipsoid.

The topographic and geoid surfaces are irregular surfaces, unlike the reference ellipsoid, which is a smooth mathematical approximation of the physical surface (geoid and/or topography).

The definition of a geodetic reference system is therefore based on these three surfaces, their mutual relations and their changes over time. The relationship between the height above the ellipsoid (the geodetic height) h, the height above sea level (orthometric height) H, and the geoid height N is illustrated in Fig. 1.4 and by the formula:

$$h = H + N. \tag{1.1}$$

[Note that in Fig. 1.4 the geoid is shown above the ellipsoid. However, in some regions, e.g. the continental United States, the geoid is actually below the ellipsoid, and therefore the value of the geoid height is negative there.]

The EGM2008 geoid model, shown in Fig. 1.5, ranges between peak values −106.6 m in the Indian Ocean south of India and 87.6 m north of Australia. The origins of these extremes are briefly discussed in Sects 8.1.1.2 and 8.2.2. Large negative geoid undulations are also seen in the Ross Sea in the Antarctic region and in NE N. America, related to postglacial mass deficiencies, and also in Mongolia, the Pacific offshore of California and the Atlantic Ocean east of the Caribbean islands.

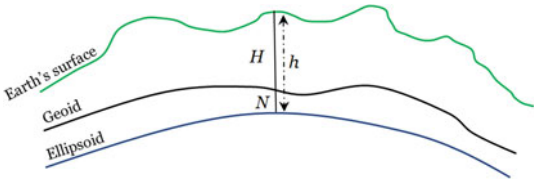

Fig. 1.4 Geodetic descriptions of the Earth's surface

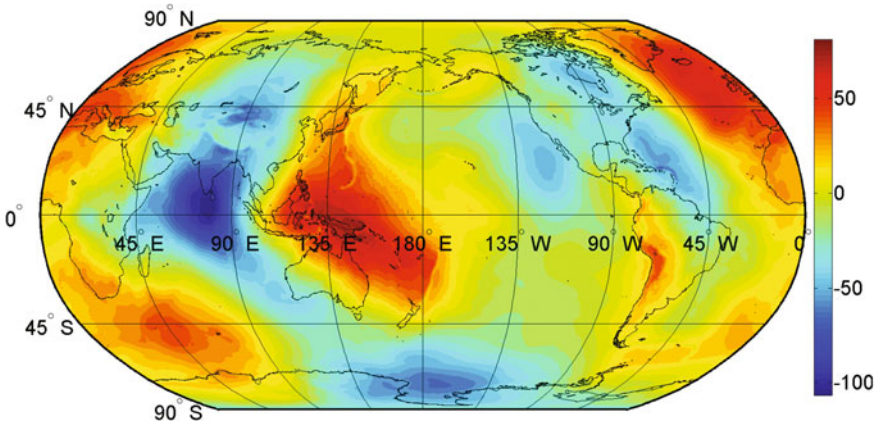

Fig. 1.5 Global geoid model obtained from EGM2008 model (up to degree 2160, corresponding grid resolution is 5′ × 5′). Unit: metre

Major positive undulations can also be found in the Andes in S. America, in N. Atlantic Ocean (sloping from both the east and west towards the mid-Atlantic ridge) and in the middle and south of Europe, as well as in SE of Africa.

1.5.2 Different Types of Reference Systems

Three-Dimensional System (3-D)

The need for a 3-D reference system has increased with the increased use of satellite methods (GNSS). An example of such a system is WGS84 by which the position of a point can be defined in a clear way all over the Earth. Figure 1.6 shows a position in 3-D expressed with Cartesian, geocentric coordinates (X, Y, Z) and geodetic coordinates [latitude (φ), longitude (λ) and ellipsoidal height (h)]. The coordinates can be converted between the two systems; see Sect. 2.3. The geocentric coordinate system places the origin at the (mass) centre of the Earth and the Z axis is directed

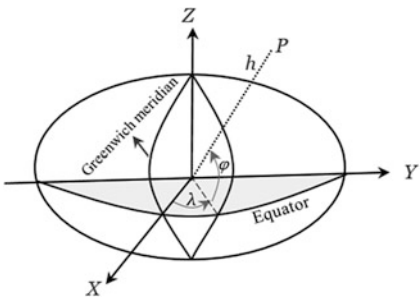

Fig. 1.6 Illustration of geocentric cartesian (X, Y, Z) and geodetic (φ, λ and h) coordinate systems

along the Earth's rotational axis. The X- and Y-axes are located in the equatorial plane with the X-axis passing through the Greenwich meridian, and the Y-axis points to the east at a right angle to the X-axis. In this way the (X, Y, Z)-system becomes a right angle system.

The geodetic coordinates latitude and longitude are used to express positions in angular dimensions. The latitude is the angle in the north-south direction with latitude zero (0) in the equatorial plane, 90° N in the North and 90° S in the South Poles. The longitude indicates the angle in the east-west direction and is based on the prime meridian, being zero along the meridian through Greenwich, counted positively eastward and westward negatively, (or with E or W for east and west, respectively).

Plane Coordinate Systems (2-D)

A 2-D reference system (or a plane coordinate system) has no direct link to the height component. Grid coordinates in the plane for a point on or above the reference ellipsoid, calculated by a selected (map) projection, is one example. The map projection depicts any curved surface or line on Earth on the mapping plane. The choice of projection is controlled by the application and projection characteristics such that the purpose of the map is met. Most geodetic maps are made in Transverse Mercator (Gauss-Krüger) projection.

Height Systems (1-D)

Height above the sea is defined in a selected height system. A height system consists of a number of fixed points (benchmarks), which are marked on the ground. In the future, the height system may be defined as a digital database versus a defined geoid model without fixed points. The height of each fixed point is carefully measured (usually by levelling). The height system is also linked to a number of parameters that define the height of the fundamental benchmark at the zero point of the system and the epoch of the height calculation that the heights are valid for, as well as a possible temporal rate of change of height of each benchmark. The fixed points are then used as starting points to measure height differences to and the heights of other objects in the surroundings. Geoid and land uplift models are closely associated with the height system. When the term "height" is used, it usually means the height above sea level, or, more precisely, the height along the plumb line above the geoid. The geoid surface is always perpendicular to the plumb line. The altitude/orthometric height (H) of an arbitrary point P is therefore the distance along the vertical/plumb line reckoned from the geoid to P. See Fig. 1.7.

Over time, height systems are affected by a number of external factors: fixed points are destroyed or moved because of subsidence or land uplift. As a result, a fixed point usually has different heights in different height systems. Differences between systems also occur when they have different fundamental/zero points. For example, Normaal Amsterdam Peil (NAP) is used as zero point in Sweden, Finland, Norway and Denmark and also in Poland, Germany and the Netherland, while many former Soviet Union countries use the Baltic Sea Datum from 1977 with zero point defined by mean sea level at Kronstadt (in 1933).

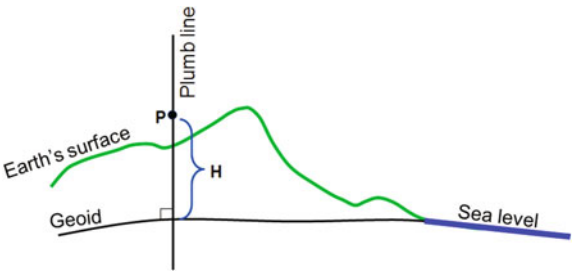

Fig. 1.7 Illustration of a height system with height *H* above the geoid

Gravity System

A reference gravity system is realized through a gravity network, which can be divided into different orders. Careful preparations for establishing a new, accurate and reliable gravity network are important, and it will help also in defining an accurate reference system.

The first international gravity network was established in 1971 by the International Union of Geodesy and Geophysics (IUGG), named the International Gravity Standardization Net (IGSN) 1970. The IGSN is a worldwide gravity network, consisting of 1854 stations with 10 absolute gravity measurements collected over the 20 years preceding 1971. Today most countries have introduced their own gravity networks and try to connect them with surrounding countries, and much of the data in Europe were delivered into the database of the Unified European Gravity Network 2002 Project (UEGN2002) (see Kenyeres et al. 2001).

Most geodetic measurement techniques are affected in one way or another by the magnitude and direction of gravity, and such information is needed, for example, in precise levelling, geoid determination and geological mapping.

1.5.3 Major Geodynamical Effects on Reference Frames

In this section, some major geodynamical phenomena that affect the reference frames will be presented. They are plate motion, glacial isostatic adjustment (postglacial rebound) and Earth tide.

Plate Motion

Continental drift, the movement of the Earth's continents relative to each other, was discovered by the German scientist Alfred Wegner in the 1930s. Nowadays, the concept of continental drift has been replaced by the concept of plate tectonics. The processes of seafloor spreading, rift valley formation and subduction are the main geologic forces underlying plate motion. Seafloor spreading occurs when molten rock rises from inside the Earth and makes new seafloor. Rift valleys are formed when a continental landmass is splitting itself apart. For example, Fig. 1.8 illustrates that the tectonic movement that all northern parts of the European plate

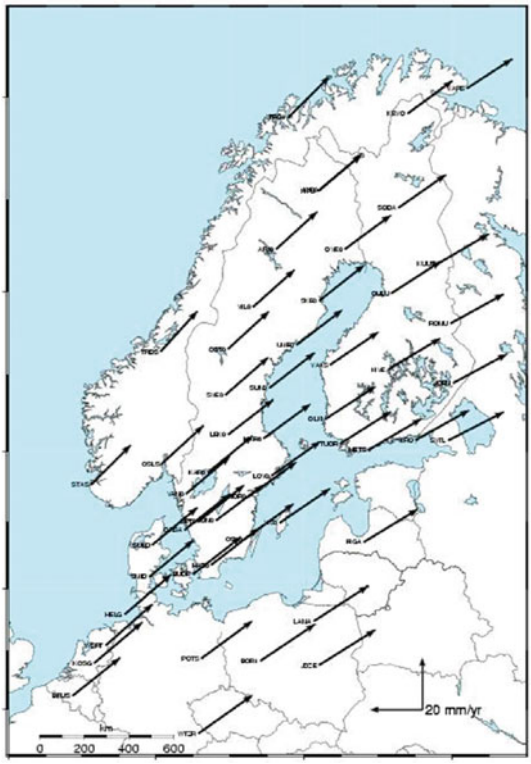

Fig. 1.8 Observed annual horizontal movement at fixed GPS reference stations in northern Europe (Reference system: ETRS89 ©Lantmäteriet, Gävle)

experience amounts to about 25 mm/yr in a NE direction. This means that, after 40 years, the coordinate values will be shifted about 1 m from a global perspective. It illustrates an essential difference between a local reference frame and a global reference frame. The local reference frame is a static system with respect to epoch related and plate tectonic movement. The global reference frame changes over time, partly by adapting to yet more recent solutions of ITRF, and through ongoing updates to the current era. (Figure 1.8 is comparable with Fig. 1.3a in the global reference system).

Post-glacial Rebound

Post-glacial rebound or Glacial Isostatic Adjustment (GIA) is the viscoelastic mass rebound of the crust from its previous depression by the ice load during the last (and previous) ice age (ages). For a long time, it was believed in the Nordic region that the sea was withdrawing. The historical evidence was based on studies of ancient shorelines that today are located on land. However, in the 1700s, it was observed instead that it is the solid crust that is rising. As proposed by Anders Celsius, watermarks were carved into rocks along the coastlines in Sweden and Finland to measure the relative sea level change. From the mid-1800s, mareographs (i.e. a type

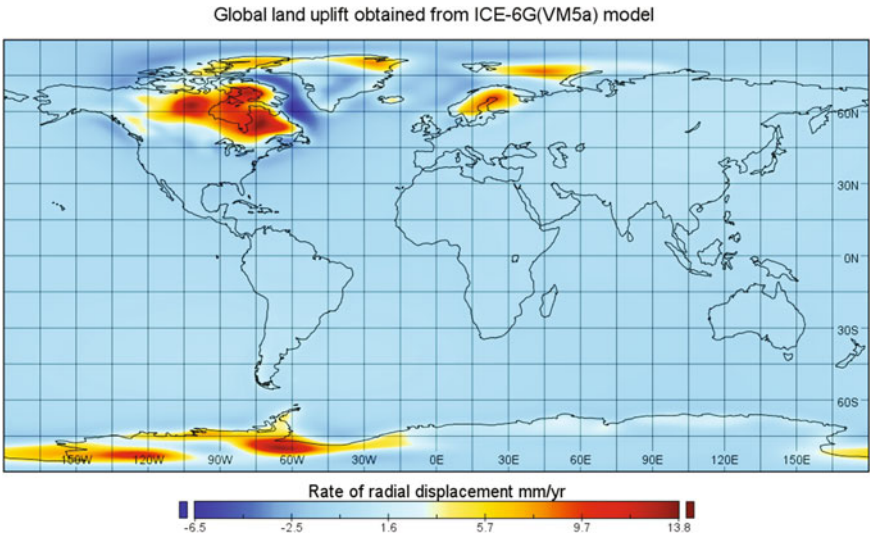

Fig. 1.9 Global land uplift model obtained based on the GIA model ICE-6G (VM5a). For details, see Peltier et al. (2015)

of accurate tide gauge) were established to more precisely monitor sea level change. Today, the relative uplift may also be determined using other types of observations, such as repeated precise levelling and GNSS observations.

Although GIA is a global phenomenon, it mostly affects the previously (or presently) glaciated regions in Laurentia, Fennoscandia, the Barents Sea, Iceland and parts of Antarctica and Greenland (see Fig. 1.9). The land uplift is of the order of 1–1.5 cm per year or less, and in the peripheral regions land is even sinking some mm/yr. The rebound also results in horizontal motions, which, in the case of Fennoscandia, ranges to about 2 mm/yr in the central uplift region in the Bothnia Bay and decreases away from this region (see Fig. 1.10). This means that geodetic projects with high accuracy, like precise levelling projects, must be corrected for the land uplift and the height systems must be time tagged.

There are various geophysical GIA models. Figure 1.9 shows a model of present-day mass change due to post-glacial rebound and the reloading of the ocean basins with seawater. The uplift rate (rate of radial displacement) is obtained using the ICE-6G (VM5a) model. In ICE-6G (VM5a), the most recently available GPS observations have been employed (Peltier et al. 2015). Red and yellow areas indicate rising due to the removal of the ice sheets. Blue areas indicate falling as mantle material moved away from these areas in order to supply the rising areas, and because of the collapse of the fore bulges around the ice sheets. Minimum and maximum values are −6.5 and 13.8 mm/yr. Uplift varies and is greatest in Canada. The next largest is in northern Sweden at the Gulf of Bothnia (about 1 cm/year) and at least in Skåne (approximately 0 cm/year).

In connection with the final calculation of the height systems in the areas with land uplift phenomena, a land uplift model should be considered to correct the height system. For example, the NKG2005LU model has been adopted as the official Nordic land uplift model of the Nordic Commission of Geodesy (NKG) in order to correct the heights. It is based on observations from mareographs, measurements from repeated precision balancing and data from fixed GPS-reference stations. For the area outside the Nordic countries, uplift is estimated from a geophysical model, which is based on the theories of the ice thickness and properties of the mantle and the crust. Ågren and Svensson (2007) presented two types of land uplift: apparent (with respect to sea level) and absolute (with respect to ellipsoid or, roughly, geoid). The relationship between apparent and absolute land uplift is given by:

$$LU_{absolute} = \left(LU_{apparent} + 1.32\right) 1.06\,\mathrm{mm/year}$$

Figure 1.10 shows the horizontal components of the global crustal motion.

Tidal Effects

The tidal acceleration at a point on Earth is the difference between the attraction of the Sun and the Moon (and planets) at the point and that at the centre of the Earth. This attraction creates both ocean and body tides, which directly affect all geodetic measurements. Most spectacular is the sea water tide, e.g. in the Bristol channel, England, and the Bay of Fundy, Canada, where it can reach 16 m. Due to the ocean tide and body tide, the Earth deforms because it is not a rigid body but behaves elastically. The tidal forces and deformations are both periodic and permanent. Most prominent are the diurnal and semi-diurnal periods, but there are also long-term periods of 8.5 and 18.6 years that can be hard to distinguish from secular motions due to other causes. The vertical displacement of the Earth surface may reach about 0.33 m, while the geoid height may change as much as 0.69 m (e.g. Vaníček et al. 1987, p. 186). There are different tidal systems, which depend on how the permanent tide is treated:

- The non-tidal system that comprises the permanent tidal effects (attraction and deformation) is completely isolated from geodetic quantities. The shape and gravity field of the Earth deviate from reality in this system.
- The mean-tidal system, which contains both the permanent tidal attraction and the permanent tidal deformation of the Earth, is retained. The shape and gravity field of the Earth correspond to reality in this system.
- The zero-tidal system where the permanent tidal attraction is eliminated but the permanent tidal deformation is retained. The gravity field of the Earth deviates from reality, and the shape of the Earth corresponds to reality in this system.

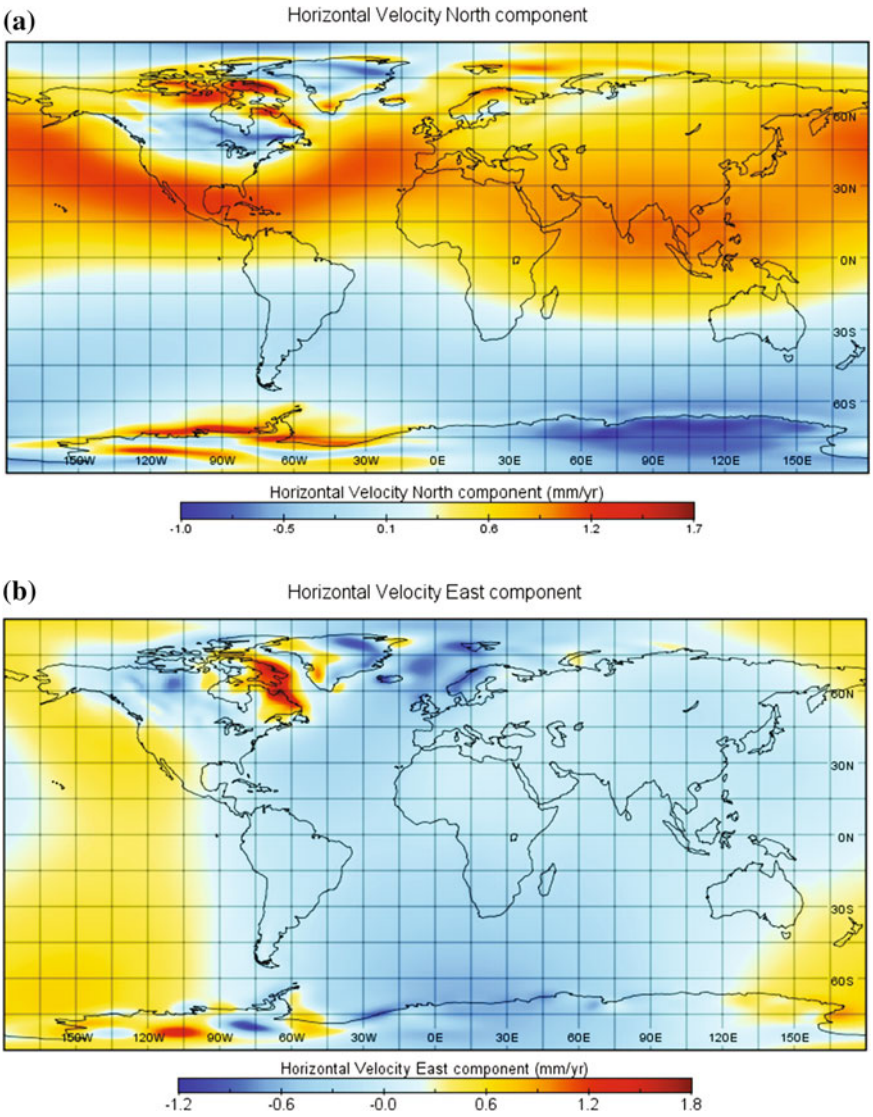

Fig. 1.10 North and east components of global crustal motion obtained based on ICE-6G (VM5a). From Peltier et al. (2015)

1.5.4 Geodetic Reference System 1980

The Geodetic Reference System 1980 (Moritz 1980b) was adopted at the XVII General Assembly of the IUGG in Canberra, Australia, in December 1979. Several of its constants are given in Table 1.1. GRS80 is (still) recommended to be used as the primary global geodetic system for various geoscience applications.

Normal Gravity

The normal gravity γ_0 can be computed at any point on the reference ellipsoid by the formula:

$$\gamma_0 = \gamma_e \frac{1 + k \sin^2 \varphi}{\left(1 - e^2 \sin^2 \varphi\right)^{1/2}} \tag{1.2}$$

where

$$k = \frac{b\gamma_p - a\gamma_e}{a\gamma_e}. \tag{1.3}$$

Table 1.1 Parameters of geodetic reference system 1980 (GRS80) (Moritz 1980b)

Conventional constants	Symbol	Value	Unit
Defining constants (exact)			
Semi-major axis of ellipsoid	a	6 378 137	m
Geocentric gravitational constant	GM	$3\ 986\ 005 \times 10^8$	m^3/s^2
Dynamic form factor	J_2	$108\ 263 \times 10^{-8}$	–
Spherical-harmonic coefficients	J_4	$-2\ 370\ 912\ 22 \times 10^{-14}$	–
Spherical-harmonic coefficients	J_6	$6\ 083\ 47 \times 10^{-14}$	–
Spherical-harmonic coefficients	J_8	-1427×10^{-14}	–
Angular velocity	ω	$7\ 292\ 115 \times 10^{-11}$	rad/s
Derived geometric constants			
Semi minor axis	b	6 356 752.3141	m
Linear eccentricity	E	521854.0097	m
First eccentricity (= e)	e^2	0.006 694 380 022 90	–
Second eccentricity (= e′)	e'^2	0.006 739 496 775 48	–
Flattening	f	0.003 352 810 681 18	–
Reciprocal flattening	f^{-1}	298.257 222 10	–
Derived physical constants			
Normal potential at ellipsoid	U_0	$6\ 263\ 686.0850 \times 10$	m^2/s^2
Normal gravity at equator	γ_e	9.780 326 7715	m/s^2
Normal gravity at pole	γ_p	9.832 186 3685	m/s^2
Gravity flattening	$(\gamma_p - \gamma_e)/\gamma_e$	$5.302440112 \times 10^{-3}$	–

For determining the normal potential and normal gravity in space above the ellipsoid, we refer to Sect. 3.2.2.

Tidal System

In 1983, the IAG adopted the zero-tidal corrections to be used in geodetic systems. This implies that the permanent tidal attraction ($-\partial\overline{W}^{tide}/\partial r$), where $\overline{W}^{tide}$ is the permanent tidal potential, is eliminated, whereas the permanent tidal deformation is retained in gravity measurements, yielding the gravity correction:

$$dg^{tide} = \delta\frac{\partial W^{tide}}{\partial r} + (1-\delta)\frac{\partial\overline{W}^{tide}}{\partial r}, \tag{1.4}$$

where W^{tide} is the (total) tidal potential and $\delta \approx 1.16$ is the gravity Love number.

Similarly, the *height above the zero-geoid* is obtained by adding the correction

$$dH^{tide} = \gamma\frac{W}{g} + (1-\gamma)\frac{\overline{W}}{g}. \tag{1.5}$$

Assuming that all gravity observations have been corrected according to Eq. (1.4), the geoid height determined by Stokes' formula will also refer to the zero-tide system.

References

Ågren J, Svensson R (2007) Postglacial land uplift model and system definition for the new swedish height system RH 2000. Reports in geodesy and geographical information systems, LMV-Rapport 2007:4, Lantmateriet, Gävle, Sweden

Altamimi Z, Collilieux X, Métivier L (2011) ITRF2008: an improved solution of the international terrestrial reference frame. J Geodesy 85(8):457–473. doi:10.1007/s00190-011-0444-4

Amante C, Eakins BW (2009) ETOPO1 1 Arc-minute global relief model: procedures, data sources and analysis. NOAA technical memorandum NESDIS NGDC-24. National Geophysical Data Centre, NOAA. doi:10.7289/V5C8276M

Bird P (2003) An updated digital model of plate boundaries. Geochem Geophys Geosyst 4 (3):1027. doi:10.1029/2001GC000252

Bjerhammar A (1962) Gravity reduction to an internal sphere. Royal Institute of Technology, Division of Geodesy, Stockholm, Sweden

Bjerhammar A (1975) Discrete approaches to the solution of the boundary value problem in physical geodesy. Boll Geo Sci Affini 34:185–240

Bjerhammar A, Stocki S, Svensson L (1980) A geodetic determination of viscosity. division of geodesy. The Royal Institute of Technology, Stockholm, Sweden

Drewes H (2009) Reference systems, reference frames, and the geodetic datum—basic considerations. In: Sideris MG (ed) Observing our changing earth, IAG Symposia, vol 133. Springer, pp 3–9. doi:10.1007/978-3-540-85426-5_1

Fowler CMR (2001) The solid earth: an introduction to global geophysics. Cambridge Press, Cambridge

Hager BH (1984) Subducted slabs and the geoid: constrains on mantle rheology rebound. J Geophys Res 89:6003–6015
Heiskanen WA, Moritz H (1967) Physical geodesy. W H Freeman and Co, San Francisco and London
Hofmann-Wellenhof B, Moritz H (2006) Physical geodesy, 2nd edn. Springer, Berlin
Hotine M (1969) Mathematical geodesy. US Department of Commerce, Washington, D.C
Jacoby W, Smilde PL (2009) Gravity interpretation—fundamentals and application of gravity inversion and geological interpretation. Springer, Heidelberg
Jekeli C, Upadhyay TN (1990) Gravity estimation from STAGE, a satellite-to-satellite tracking mission. J Geophys Res 95(B7):10973–10985. doi:10.1029/JB095iB07p10973
Kaula WM (1966) Theory of satellite geodesy. Blaisdell Publishing Company, London
Kaula WM (1972) Global gravity and mantle convection. Tectonophysics 131(4):341–359 (In: Ritsema AR (ed) The upper mantle)
Kenyeres A, Boedecker G, Francis O (2001) Unified European Gravity Reference Network 2002 (UEGN2002): a status report. International Gravity and Geoid Commission, Thessaloniki, Greece
Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Geod Inst Copenhagen Medd No. 44. Also in Borre K (ed) Mathematical foundation of geodesy—selected papers of T Krarup, Springer, 2006
Lambeck K (1988) Geophysical geodesy. The slow deformations of the earth. Clarendon Oxford University Press, New York, xii, 718 pp
Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture notes in earth sciences. Springer, New York
Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitational field and figure of the earth. Translated from Russian (1960) Israel program for scientific translations. Jerusalem, Israel
Moritz H (1980a) Advanced physical geodesy. Herbert Wichmann Verlag, Karlsruhe
Moritz H (1980b) Geodetic reference system 1980. Bull Géodésique 54(3):395–405. doi:10.1007/BF02521480
Moritz H (1990) The figure of the earth. Wichmann, Karlsruhe
Pavlis N, Factor K, Holmes SA (2007) Terrain-related gravimetric quantities computed for the next EGM. Presented at the 1st international symposium of the international gravity service 2006, Aug 28–Sept 1, Istanbul, Turkey
Pavlis NA, Simon AH, Kenyon SC, Factor JK (2012) The development and evaluation of the earth gravitational model 2008 EGM2008. JGR 117:B04406
Peltier WR, Argus DF, Drummond R (2015) Space geodesy constrains ice-age terminal deglaciation: the global ICE-6G_C VM5a model. J Geophys Res Solid Earth 120:450–487. doi:10.1002/2014JB011176
Richards MA, Hager BH (1984) Geoid anomalies in a dynamic earth. J Geophys Res 89:5987–6002
Runcorn S (1967) Flow in the mantle inferred from the low degree harmonics of the geopotential geophys. J Int 14(1–4):375–384. doi:10.1111/j.1365-246X.1967.tb06253.x
Sansó F, Sideris MG (eds) (2013) Geoid determination—theory and methods. Lecture notes in earth system sciences, Springer, Heidelberg
Sjöberg LE (1982) On the recovery of geopotential coefficients using satellite-to-satellite range-rate data on a sphere. Bull Geod 56:27–39
Steffen H, Wu P (2011) Glacial isostatic adjustment in Fennoscandia—a review of data and modeling. J Geodyn 52(3–4):169–204. ISSN 0264-3707. doi:10.1016/j.jog.2011.03.002
Turcotte DL, Schubert G (2002) Geodynamics. 2nd ed. Cambridge University Press, Cambridge. xv + 456 pp. ISBN 0521 66186 2; 0 521 66624 4 (pb) (Geol Mag 139)
United Nations (2015) General assembly resolution on a global geodetic reference frame. Available from: http://ggim.un.org/knowledgebase/Attachment157.aspx. Accessed 3 June 2015

Vaníček P, Christou Nikolaos T (1994) Geoid and its geophysical interpretations. CRC Press, USA. ISBN 0-8493-4227-9

Vaníček P et al (1987) Four-dimensional geodetic positioning—report of the IAG SSG 4.96. Manuscr Geod 12(3):147–222

Watts AB (2001) Isostasy and flexure of the lithosphere. Cambridge University Press, Cambridge, xix + 458 pp

Wolff M (1969) Measurements of the Earth's gravitational potential using a satellite pair. J Geohys Res 74:5295–5300

Chapter 2
Basic Mathematics

Abstract The basic mathematics useful for this book is divided into discrete least squares theory, collocation, coordinate systems, Legendre's polynomials, spherical and ellipsoidal harmonics, the fundamentals of potential theory and regularization. Most numerical applications are based on linear least squares theory, either in the spatial domain (mainly for local studies) or by spherical harmonics in regional and global applications. For example, linear regression analysis, discrete and continuous least squares collocation are described. As problems in geodesy and geophysics are frequently non-linear, the linearization of such a problem is also presented. After introducing Legendre's polynomials and spherical harmonics, the latter type of series is used for spectral smoothing and combining sets of data. The gravitational potential on and outside the ellipsoid is also presented in ellipsoidal harmonics. One section is devoted to the basics of potential theory, including some basic concepts, Newton's integral for the potential, Laplace's and Poisson's equations and Gauss' and Green's formulas, as a well as basic boundary value problems, as a background for the rest of the book. Considering that most problems related with gravity inversion are inverse problems, regularization is needed to reach a practical solution. Hence, various approaches to regularization of solutions to inverse problems are shortly described and compared.

Keywords Basic mathematics · Collocation · Coordinate systems · Least squares theory · Legendre's polynomials · Spherical harmonics · Potential theory · Regularization

2.1 Least Squares Adjustment Theory

Least squares treatment of large data sets is common in geodesy, surveying and geophysics. Least squares collocation, widely used in geodesy, is closely related with kriging, frequently applied in geophysical prospecting. This book will apply

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_2

least squares in various ways in physical geodesy and geophysics, and the basics are provided in this section.

Definition 2.1 Let x be a stochastic estimator for the parameter μ. If the stochastic expectation $E\{x\} = \mu$ holds, we say that x is an unbiased estimator of μ and its variance is given by:

$$\sigma_x^2 = E\left\{(x-\mu)^2\right\}. \tag{2.1}$$

Definition 2.2 If $E\{x\} \neq \mu$, then x is a *biased* estimator of μ, and its Mean Square Error,

$$MSE\{x\} = E\left\{(x-\mu)^2\right\} = \sigma_x^2 + bias_x^2, \tag{2.2a}$$

is the sum of its variance given by (2.1) and the bias squared, the bias given by:

$$bias_x = E\{x\} - \mu. \tag{2.2b}$$

Equation (2.2a) follows directly from the relation:

$$MSE\{x\} = E\left\{(x-\mu)^2\right\} = E\left\{(x - E\{x\} + E\{x\} - \mu)^2\right\}. \tag{2.3}$$

2.1.1 Adjustment by Elements

Let us assume that $\mathbf{L}$ is a vector of n observations with a random error vector $\boldsymbol{\varepsilon}$. If $\mathbf{L}$ is related with the unknown parameter vector $\mathbf{X}$ (with $m < n$ elements) by the linear matrix equation (system of observation equations)

$$\mathbf{AX} = \mathbf{L} - \boldsymbol{\varepsilon}, \tag{2.4}$$

where $\mathbf{A}$ is called the design matrix, assumed to be of full rank, the least squares solution to (2.4), minimizing the weighted sum of squares of errors, $\boldsymbol{\varepsilon}^T\mathbf{P}\boldsymbol{\varepsilon}$, is:

$$\widehat{\mathbf{X}} = \left(\mathbf{A}^T\mathbf{P}\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{P}\mathbf{L}. \tag{2.5a}$$

The error vector and covariance matrix of the unknowns become:

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{L} - \mathbf{A}\widehat{\mathbf{X}} \tag{2.5b}$$

and

$$\mathbf{Q}_{\mathbf{X}} = \sigma_0^2 \left(\mathbf{A}^T \mathbf{P} \mathbf{A}\right)^{-1}. \tag{2.5c}$$

Here $\mathbf{P}$ is a positive definite weight matrix among the observations, $(\,)^{-1}$ is the inverse of the matrix in the bracket and σ_0^2 is *the variance of unit weight*. The latter can be unbiasedly estimated by:

$$s^2 = \hat{\boldsymbol{\varepsilon}}^T \mathbf{P} \hat{\boldsymbol{\varepsilon}} / (n-m) = \mathbf{L}^T \mathbf{P} \left(\mathbf{L} - \mathbf{A}\widehat{\mathbf{X}}\right) / (n-m), \tag{2.5d}$$

where n and m are the numbers of observations and unknown parameters, respectively.

Example 1 Consider a linear regression in time (t_i) with observation equations

$$a + bt_i = l_i - \varepsilon_i; \quad i = 1, \ldots, n, \tag{2.6a}$$

or, in a matrix equation

$$\begin{bmatrix} 1 & t_1 \\ .. & .. \\ 1 & t_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} l_1 \\ .. \\ l_n \end{bmatrix} - \boldsymbol{\varepsilon} \tag{2.6b}$$

with the least squares solution for the parameters a and b

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n} l_i \\ \sum_{i=1}^{n} t_i l_i \end{bmatrix} \tag{2.6c}$$

If one substitutes t_i by $\Delta t_i = t_i - t_0$, where $t_0 = \left(\sum_{i=1}^{n} t_i\right)/n$ is the mean of the observation times, one obtains:

$$\sum_{i=1}^{n} \Delta t_i = 0,$$

implying that the off-diagonal elements of the normal matrix $\mathbf{A}^T\mathbf{A}$ vanish, yielding a diagonal matrix, and the above solution is simplified to:

$$\hat{a} = \sum_{i=1}^{n} l_i / n \quad \text{and} \quad \hat{b} = \sum_{i=1}^{n} \Delta t_i l_i / \sum_{i=1}^{n} (\Delta t_1)^2 \tag{2.6d}$$

with standard errors

$$s_{\hat{a}} = s/\sqrt{n} \quad \text{and} \quad s_{\hat{b}} = s/\sqrt{\sum_{i=1}^{n} (\Delta t_1)^2}, \tag{2.6e}$$

where the variance of unit weight (s) is given by Eq. (2.5d).

This solution is useful in estimating the secular change/trend parameter b of the set of observations (l) observed at different epochs. In particular, if the time interval (Δt) between successive epochs is constant, it follows that:

$$\hat{b} = \frac{2}{\Delta t} \frac{\sum_{i=1}^{n} (2i - n - 1) l_i}{\sum_{i=1}^{n} (2i - n - 1)^2} \quad \text{and} \quad s_b = \frac{s}{\Delta t \sqrt{\sum_{i=1}^{n} (2i - n - 1)^2}}. \tag{2.6f}$$

One may also eliminate the constant a in Eq. (2.6a) by substituting each observation l_i by $l_i - \bar{l}$, where $\bar{l}$ is the mean of the observations.

The estimated secular trend $(\hat{b})$ may be sensitive to periodic signals not included in the regression analysis, in particular for long-periodic terms. If the periods (T_j) are known, their causes can be included in the adjustment by the revised formula

$$a + bt_i + \sum_{j=1}^{J} \left(c_j \cos \omega_j t_i + d_j \sin \omega_j t_i \right) = l_i - \varepsilon_i; \quad i = 1, \ldots, n, \tag{2.6g}$$

where $\omega_j = 2\pi/T_j$ and J is the number of periodic signals included in the adjustment.

If the set of observations are evenly distributed over the period and includes one or a multiple of periods, the effect of the periodic term is eliminated. More generally, the regression formula is extended to multiple regression analysis by including several types of correlated observables. Then the normal matrix $\mathbf{A}^T\mathbf{PA}$ of the solution (2.5a) will be a full matrix, and, for example, the simple solution for the trend parameter in (2.6f) does not hold, implying that the unknown parameters are correlated and mutually affect the solutions of the individual parameters.

– *Special Cases:*

- Frequently the underlying function $f_i(\mathbf{X})$ for the observation l_i is *non-linear*:

$$l_i - \varepsilon_i = f_i(\mathbf{X}); \quad i = 1, \ldots, n, \tag{2.7a}$$

and by Taylor expansion of all observation equations the system may be linearized to the matrix equation

$$\mathbf{A}\Delta\mathbf{X} = \Delta\mathbf{L} - \boldsymbol{\varepsilon}, \tag{2.7b}$$

where $\mathbf{A}$ is the design matrix as above, $\Delta\mathbf{X}$ is the (unknown) improvement of $\mathbf{X}$ versus the approximate value $\mathbf{X_0}$ and

$$\Delta\mathbf{L} = \mathbf{L} - [f_1(\mathbf{X_0}), \ldots, f_n(\mathbf{X_0})]^T \tag{2.7c}$$

Here $\mathbf{L}$ is the vector of observations l_i. By solving (2.7b) as in (2.5a), the least squares solution is achieved. As the original equations are non-linear, the solution may need to be iterated for convergence.

- If there are a priori information $\mathbf{X}^-$ of the unknown vector $\mathbf{X}$ with covariance matrix $\mathbf{Q}_X$, the matrix equation (2.4) can be augmented by the equation

$$\mathbf{IX} = \mathbf{X}^- - \boldsymbol{\varepsilon}_x; \quad E\left(\boldsymbol{\varepsilon}_\mathbf{x}\boldsymbol{\varepsilon}_\mathbf{x}^\mathbf{T}\right) = \mathbf{Q}_X, \tag{2.8a}$$

and assuming also that the observations $\mathbf{X}^-$ and $\mathbf{L}$ are uncorrelated, the improved least squares solution becomes (e.g. Sjöberg 2013, Sect. 12.1):

$$\widehat{\mathbf{X}} = \mathbf{N}^{-1}\left(\mathbf{Q}_\mathbf{X}^{-1}\mathbf{X}^- + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{L}\right) \tag{2.8b}$$

or

$$\widehat{\mathbf{X}} = \mathbf{X}^- + \mathbf{K}(\mathbf{L} - \mathbf{A}\mathbf{X}^-), \tag{2.8c}$$

where:

$$\mathbf{N} = \mathbf{Q}_\mathbf{X}^{-1} + \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A} \quad \text{and} \quad \mathbf{K} = \mathbf{Q}_\mathbf{X}\mathbf{A}^T\left(\mathbf{A}\mathbf{Q}_\mathbf{X}\mathbf{A}^T + \mathbf{Q}\right)^{-1}, \tag{2.8d}$$

and the covariance matrix of the solution vector can be written:

$$\mathbf{Q}_{\widehat{\mathbf{X}}} = \mathbf{N}^{-1} = \mathbf{Q}_\mathbf{X} - \mathbf{K}\mathbf{A}\mathbf{Q}_\mathbf{X}. \tag{2.8e}$$

It is obvious that this solution is both more stable and precise than the original solution (2.5a).
As an example, the solution of Eqs. (2.8c) and (2.8e) is useful in gravity inversion when a preliminary model of Earth parameters $(\mathbf{X}^-)$ are at hand from a previous analysis, and new data $(\mathbf{L})$ are available to improve the model. In the case that the new observation equation is non-linear, it can be linearized as described above. Several other ways of adjusting the non-linear equations in combination with the preliminary model are discussed at length in Tarantola (1987).

- Closely related adjustment schemes are condition adjustment and condition adjustment with unknowns. See, e.g. Bjerhammar (1973, Chaps. 12 and 20),

A discrete least squares problem may be *rank deficient* or *ill-conditioned*, leading to proper and numerical singularities, respectively, in the systems of equations. In the first case, there is no unique solution. In the second case, a unique solution may exist, but the system is badly conditioned such that the numerical solution may fail or be badly contaminated by errors. In solving geoscience problems by gravity inversion, the systems of equations are frequently ill-conditioned, as such problems are typically *ill-posed*, which can be handled by some type of regularization (see Sect. 2.8).

2.2 Least Squares Collocation

2.2.1 Discrete Collocation

Least Squares Collocation (LSC) is a type of interpolation and/or prediction of stochastic variables, either within one type of observable or from the observations of one type to another. In addition, the covariances among the observables as well as between these and the predicted variable are assumed to be known.

Let x and y be stochastic variables with expectations zero. The variable y is assumed to be estimated (predicted) from observations of x. The auto-covariance and error covariance matrices ($\mathbf{C}$ and $\mathbf{D}$) among the observations in the observation matrix $\mathbf{X}$, as well as the cross-covariance vector $\mathbf{c}$ between y and $\mathbf{X}$, are known. Moreover, if the variance of y, σ_y^2, is known, the prediction variance can also be estimated. In this case y can be optimally estimated/predicted in a least squares sense by the formula (Moritz 1980, Part B)

$$\hat{y} = \mathbf{c}^T(\mathbf{C}+\mathbf{D})^{-1}\mathbf{X}, \tag{2.9a}$$

and the prediction variance becomes:

$$\sigma_{\hat{y}}^2 = \sigma_y^2 - \mathbf{c}^T(\mathbf{C}+\mathbf{D})^{-1}\mathbf{c}. \tag{2.9b}$$

Proof Consider the general linear estimator

$$\tilde{y} = \mathbf{a}^T\mathbf{X}, \tag{2.10a}$$

where $\mathbf{a}$ is an arbitrary vector and the error of $\mathbf{X}$ is $\boldsymbol{\varepsilon}$ with expectation zero. Then the prediction error becomes:

$$\varepsilon_{\tilde{y}} = \tilde{y} - y = \mathbf{a}^T\mathbf{X} - y, \tag{2.10b}$$

and by assuming that $\boldsymbol{\varepsilon}$ and y are uncorrelated, the prediction variance becomes:

$$\begin{aligned}\sigma_{\tilde{y}}^2 = E\left\{\varepsilon_{\tilde{y}}^2\right\} &= \sigma_y^2 + \mathbf{a}^T E\left\{(\mathbf{X}+\boldsymbol{\varepsilon})(\mathbf{X}+\boldsymbol{\varepsilon})^T\right\}\mathbf{a} - 2\mathbf{a}^T E\{(\mathbf{X}+\boldsymbol{\varepsilon})y\} \\ &= \sigma_y^2 + \mathbf{a}^T(\mathbf{C}+\mathbf{D})\mathbf{a} - 2\mathbf{a}^T\mathbf{c} = \sigma_y^2 - \mathbf{c}^T(\mathbf{C}+\mathbf{D})^{-1}\mathbf{c} \\ &\quad + \left[\mathbf{a} - (\mathbf{C}+\mathbf{D})^{-1}\mathbf{c}\right]^T(\mathbf{C}+\mathbf{D})\left[\mathbf{a} - (\mathbf{C}+\mathbf{D})^{-1}\mathbf{c}\right] \\ &\geq \sigma_y^2 - \mathbf{c}^T(\mathbf{C}+\mathbf{D})^{-1}\mathbf{c},\end{aligned} \tag{2.10c}$$

where $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$, $\mathbf{D} = E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\}$ and $\mathbf{c} = E\{\mathbf{x}y\}$. This shows that the optimum predictor is provided for $\mathbf{a} = (\mathbf{C}+\mathbf{D})^{-1}\mathbf{c}$, and the predictor and its variance follow from (2.9a, b).

If the expectation of x does not vanish, collocation can still be applied as above after first removing the bias, trend or systematic error by a least squares adjustment by elements as above. The whole procedure, including trend removal, is the general form of least squares collocation (Moritz 1980, Parts B and C).

When collocation is applied to the gravity field of the Earth, it is unrealistic to assume that the signal (but, on the contrary, the error) is stochastic, and this can only be assumed as an approximate model. As a result, also the signal covariance functions needed are in doubt. In the application of collocation the statistical expectation operator is replaced by a space average operator, which is not trivial. Lauritzen (1973) even proved that a Gaussian stochastic process of gravity is not ergodic, implying that "even if we knew gravity all over the Earth, we would not be able to find the true covariance function". If there is only one realization available of the stochastic process, we cannot determine the true covariance function.

Nevertheless, discrete collocation is an important method for inter- and extrapolation that also provides an approximate estimate of the prediction variance. Even if this method is not the most precise technique, it is frequently also used for determining various quantities from gravity data. However, one should also bear in mind that collocation leads to large matrix systems when many observations are at hand.

In geology and geophysics, a similar concept named kriging has been developed for applications in geostatistics (see, e.g. Matheron 1963). The essential difference to collocation is the variogram that replaces the covariance function used in collocation. See, e.g. Dermanis (1984) for further details.

2.2.2 Continuous Collocation

Let x be a continuous stochastic process on the sphere with expectation zero and auto-covariance function $c_{xx}(P, Q)$. Its observation $\bar{x}$ is contaminated by the error ε, which is assumed to be uncorrelated with the true signal x, has expectation 0 and auto-covariance function $d(P, Q)$.

Problem It is requested to determine the least squares solution to the linear continuous estimator on the sphere:

$$\tilde{y}_P = \iint_\sigma h\bar{x}d\sigma, \tag{2.11}$$

where h is an unknown kernel function to be determined such that the variance of $\tilde{y}$ is a minimum. Here σ is the unit sphere. In addition, the cross-covariance function between the signals y and x, i.e. $c_{yx}(P,Q) = c_{xy}(P,Q)$, is assumed to be known.

Solution (Sjöberg 1979): The error of $\tilde{y}$ is:

$$\varepsilon_y = \tilde{y} - y = \iint_\sigma h\bar{x}d\sigma - y, \tag{2.12}$$

with the prediction variance

$$\sigma^2_{\tilde{y}}(P) = E\left\{(\tilde{y}-y)^2\right\} = \sigma^2_y(P) - 2\iint_\sigma h(P,Q)c_{yx}(P,Q)d\sigma_Q + \iint_\sigma h(P,Q)\left[\iint_\sigma h(P,Q')\{c_{uu}(Q,Q' + d((Q,Q')\}d\sigma_{Q'}\right]d\sigma_Q\,. \tag{2.13}$$

It follows that the minimum prediction variance

$$\sigma^2_{\tilde{y}}(P) = \sigma^2_y(P) - 2\iint_\sigma \hat{h}(P,Q)c_{yx}(P,Q)d\sigma_Q \tag{2.14}$$

is obtained by:

$$c_{yx}(P,Q) = \iint_\sigma \hat{h}(P,Q')\{c_{uu}(Q,Q') + d(Q,Q')\}d\sigma_{Q'}. \tag{2.15}$$

which is the so-called Wiener-Hopf integral equation for the kernel function h. The least squares solution is thus given by Eq. (2.11) with h given by Eq. (2.15).

If there are two different sets of stochastic processes, x_1 and x_2 on the sphere, observed with random errors ε_1 and ε_2, and all covariance functions are known (with obvious notations similar to the above example), a related stochastic process y can be optimally estimated from the general combined estimator

$$\tilde{y}_P = \iint_\sigma (h_1(P,Q)\bar{x}_1(Q) + h_2(P,Q)\bar{x}_2(Q))d\sigma. \tag{2.16}$$

The least squares solution for functions h_1 and h_2 are the solutions to the integral equations

$$c_{yx_1}(P,Q) = \iint_\sigma \left[\hat{h}_1(P,Q')\bar{c}_{11}(Q',Q) + \hat{h}_2(P,Q')c_{12}(Q',Q)\right]d\sigma' \tag{2.17a}$$

and

$$c_{yx_2}(P,Q) = \iint_\sigma \left[\hat{h}_2(P,Q')\bar{c}_{22}(Q',Q) + \hat{h}_1(P,Q')c_{21}(Q',Q)\right]d\sigma', \tag{2.17b}$$

where:

$\bar{c}_{11} = c_{11} + d_{11}$ and $\bar{c}_{22} = c_{22} + d_{22}$; d_{ii} are the respective noise covariance functions.

The expected least squares prediction variance becomes:

$$\sigma^2_{\hat{y}}(P) = \sigma^2_y(P) - \iint_\sigma \left[\hat{h}_1(\mathrm{P},\mathrm{Q}) \iint_\sigma \{\bar{\mathrm{c}}_{11}(\mathrm{Q},\mathrm{Q}')\hat{h}_1(\mathrm{P},\mathrm{Q}) + 2\mathrm{c}_{12}(Q,Q')\hat{h}_2(P,Q')\}d\sigma'\right]d\sigma$$
$$- \iint_\sigma \left[\hat{h}_2(\mathrm{P},\mathrm{Q}) \iint_\sigma \{\bar{\mathrm{c}}_{22}(\mathrm{Q},\mathrm{Q}')\hat{h}_2(\mathrm{P},\mathrm{Q}')\}d\sigma'\right]d\sigma. \tag{2.18}$$

The solutions to the kernel functions h_i can be conveniently determined from Eqs. (2.17a, b) by expressing all functions in spherical harmonics (see Sect. 2.8.2). If the stochastic processes are harmonic in the exterior of the sphere (which is the case if they are related with the gravity field and there are no topographic and atmospheric masses outside the sphere), the predictions $\hat{y}_p$ can be extended to any point on or outside the sphere (Sjöberg 1979). The Wiener filter is further discussed in Sect. 2.8.2 as a method of *regularization*.

2.3 Coordinate Systems

Consider the point P in an Earth-fixed Cartesian coordinate system (x, y, z), with origin at the Earth's centre of gravity, the z-axis along the Earth's axis of rotation and the (x, y)-plane in the equatorial plane with the x-axis in Greenwich meridian and y-axis at right angle to the east.

Then the Cartesian coordinates can also be expressed in spherical coordinates (see Fig. 2.1)

$$\begin{aligned} x &= r\cos\psi\cos\lambda \\ y &= r\cos\psi\sin\lambda \\ z &= r\sin\psi, \end{aligned} \tag{2.19}$$

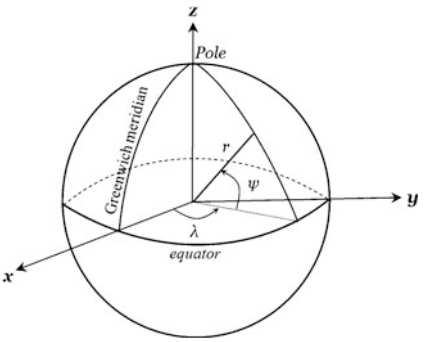

Fig. 2.1 Relation between Cartesian and spherical coordinates

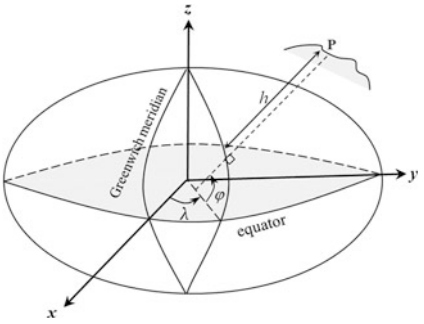

Fig. 2.2 Relation between Cartesian and geodetic coordinates

where r and ψ are the geocentric radius and latitude, respectively, and λ is the longitude.

The geodetic coordinates (φ, λ, h) = (latitude, longitude, height) are related to the Cartesian coordinates by the formulas (see Fig. 2.2)

$$\begin{aligned} x &= (\mathbb{N}+h)\,\cos\varphi\,\cos\lambda \\ y &= (\mathbb{N}+h)\,\cos\varphi\,\sin\lambda \\ z &= \left[\mathbb{N}(1-\mathrm{e}^2)+h\right]\,\sin\varphi, \end{aligned} \tag{2.20}$$

where:

$$\mathbb{N} = a/\sqrt{1-e^2\sin^2\varphi} \tag{2.21}$$

is the radius of curvature in the prime vertical of the reference ellipsoid with semi-major and semi-minor axes a and b, and $e^2 = (a^2 - b^2)/a^2$ is the eccentricity of the ellipsoid squared.

Next we present the relationship between the Cartesian and ellipsoidal coordinates (u, β, λ) (see Fig. 2.3)

$$\begin{aligned} x &= \sqrt{u^2 + E^2}\, \cos\beta\, \cos\lambda \\ y &= \sqrt{u^2 + E^2}\, \cos\beta\, \sin\lambda \\ z &= u\, \sin\beta, \end{aligned} \tag{2.22}$$

where:

$$E = ae = \sqrt{a^2 - b^2}$$

is the linear eccentricity.

By considering a point on the reference ellipsoid (i.e., with $h = 0$ and $u = b$), it follows from the above equations that geocentric, geodetic and reduced latitudes (ψ, φ, β) are related by the equations

$$\frac{z}{\sqrt{x^2 + y^2}} = \tan\psi = \left(1 - e^2\right) \tan\varphi = \sqrt{1 - e^2}\, \tan\beta. \tag{2.23}$$

The Cartesian coordinates can be determined from the curvilinear coordinates above by straightforward transformations. The inverse transformations are much more cumbersome to derive, unless the computational point is located on the reference ellipsoid. An exception is the longitude, which for all three curvilinear coordinate systems above is the same and can be determined by:

$$\lambda = 2 \arctan \frac{y}{x + \sqrt{x^2 + y^2}}. \tag{2.24}$$

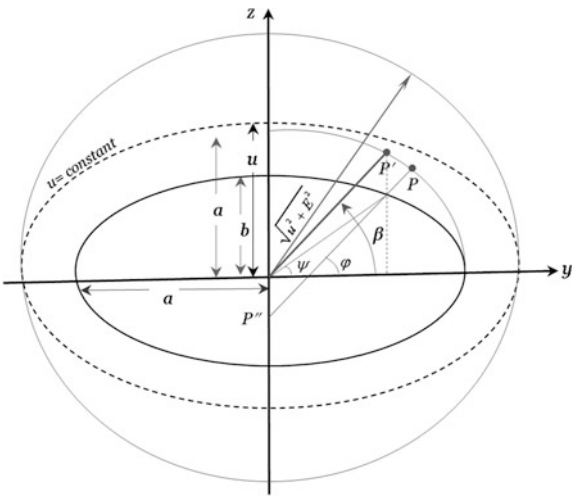

Fig. 2.3 Relations between Cartesian coordinates and geocentric, geodetic and reduced latitudes

Equation (2.19) can be inverted to provide the spherical coordinates (r, ψ) from the Cartesian ones by:

$$\tan\psi = \frac{z}{p} \quad \text{and} \quad r = \sqrt{p^2 + z^2}, \tag{2.25a}$$

where:

$$p = \sqrt{x^2 + y^2}. \tag{2.25b}$$

From Eq. (2.22) one obtains the inverse transformation from Cartesian coordinates to ellipsoidal coordinates by:

$$u = \sqrt{\left(w^2 + \sqrt{w^4 + 4E^2z^2}\right)/2} \tag{2.26a}$$

and

$$\tan\beta = \frac{z}{p}\sqrt{1 + (E/u)^2}, \tag{2.26b}$$

where:

$$w^2 = p^2 + z^2 - E^2. \tag{2.26c}$$

There are numerous solutions in the literature to the geodetic latitude φ and height h; some solutions are iterative (e.g. Heiskanen and Moritz 1967, p. 183; Fukushima 2006; others are approximate, (e.g. Hoffmann-Wellenhof et al. 2008, p. 280):

$$\varphi = \arctan\frac{z + (e')^2 b \sin^3 o}{p - e^2\cos^3 o}, \tag{2.27a}$$

where the auxiliary argument o is given by:

$$\tan o = z/\left(p\sqrt{1 - e^2}\right) \quad \text{and} \quad (e')^2 = \left(a^2 - b^2\right)/b^2, \tag{2.27b}$$

and there are also several exact solutions (see below).

Two efficient iterative methods were presented by Fukushima (2006) by applying Halley's third-order method to solve non-linear equations (Danby 1988). This method can be derived as follows. By squaring and adding the first two of Eq. (2.20), one obtains:

$$p = (\mathbb{N} + h)\cos\varphi, \tag{2.28}$$

and from this equation, and the last equation of (2.20) h can be eliminated, resulting in an equation in φ alone:

$$f(\varphi) = p\tan\varphi - z - e^2 a\frac{\sin\varphi}{\sqrt{1-e^2\sin^2\varphi}} = 0, \tag{2.29a}$$

and by taking advantage of Eq. (2.23), one finally obtains:

$$F(T) = PT - Z - e^2\frac{T}{\sqrt{1+T^2}} = 0, \tag{2.29b}$$

where $P = p/a$, $Z = z/a$ and $T = \tan\beta$.

Applying Halley's third-order method, the iteration for T uses the formula

$$T_{n+1} = T_n - \frac{F(T_n)}{F'(T_n) - F''(T_n)F(T_n)/(2F'(T_n))}, \tag{2.30a}$$

where n is the iteration step and the first- and second-order derivatives are:

$$F'(T_n) = P - e^2/\left(1+T_n^2\right)^{3/2} \quad \text{and} \quad F''(T_n) = 3e^2T_n/\left(1+T_n^2\right)^{5/2}. \tag{2.30b}$$

As the second-order derivative is rather cumbersome, it may be neglected, yielding Newton's second-order iteration formula

$$T_{n+1} = T_n - \frac{F(T_n)}{F'(T_n)}. \tag{2.31}$$

A suitable start value for T_n could be $T_0 = Z/[P(1-e^2)]$.

Fukushima (2006) showed that this iterative method is faster than any of the other methods published for the transformation of Cartesian to geodetic coordinates at an accuracy within $6" \times 10^{-6}$ for elevations ranging between -10 and 30,000 km.

Note that the iteration using the latitude as an argument will have a problem for high latitudes. In such situations, Sjöberg (1999) proposed iteration by using the co-latitude. Another way to circumvent the problem close to the pole is to substitute the unknown latitude φ by: $\phi = \varphi \mp \pi/4$.

An exact solution of geodetic height and latitude (Sjöberg 2008; slightly revised) reads as follows.

By introducing the new unknown $k = (\mathbb{N}+h)/\mathbb{N}$ one obtains the following equations from Eqs.(2.20), (2.21) and (2.23):

$$P = \left(x^2+y^2\right)/a^2 = k^2\cos^2\beta \tag{2.32a}$$

and

$$Q = z^2\left(1 - e^2\right)/a^2 = \left(k - e^2\right)\sin^2\beta, \tag{2.32b}$$

and by combining these two equations such that β is eliminated, one arrives at a fourth-order equation in k:

$$\frac{P}{k^2} + \frac{Q}{(k - e^2)} = 1. \tag{2.32c}$$

Using the notations

$$r = \left(P + Q - e^2\right)/6 \quad \text{and} \quad s = PQe^4/(4r^3) \tag{2.33}$$

followed by

$$\begin{aligned} t &= \sqrt[3]{1 + s + \sqrt{2s + s^2}},\ u = r\left(1 + t + t^{-1}\right),\ v = \sqrt{u^2 + e^4P} \\ &\text{and} \quad w = e^2\frac{v + u - P}{2v}, \end{aligned} \tag{2.34a}$$

the only real and positive solution for k becomes

$$k = w + \sqrt{w^2 + v + u} \quad \text{and also} \quad W = k\sqrt{\frac{1 - e^2}{k^2 - e^2P}}. \tag{2.34b}$$

Inserting (2.21) into the definition of k above (2.32a) the geodetic height can now be determined by

$$h = \frac{a}{W}(k - 1). \tag{2.35}$$

If P is small vs. Q then the latitude is given by Eq. (2.32a) as

$$\varphi = \pm\arccos(W\sqrt{P}/k)(\text{with the same sign as z}). \tag{2.36a}$$

Otherwise Eq. (2.32b) yields

$$\varphi = \pm\arcsin\left(\frac{k}{k - e^2}\sqrt{\frac{Q}{k^2 - e^2P}}\right)(\text{with the same sign as z}). \tag{2.36b}$$

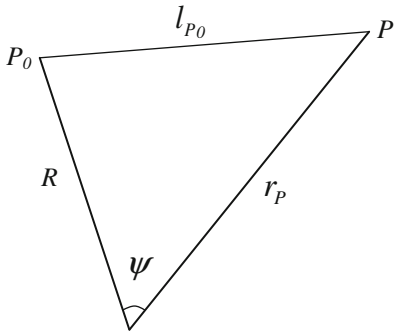

Fig. 2.4 Spherical distance/geocentric angle ψ between the radius vectors for points P_0 and P

2.4 Legendre's Polynomials

Legendre's polynomials and spherical harmonics are important in regional and global data representations. The generating function for Legendre polynomials is the inverse distance between two points P_0 and P, located on a sphere of radius R and outside the sphere at distance r_P from the centre of the sphere, respectively:

$$l_{P0}^{-1} = \frac{1}{\sqrt{r_P^2 + R^2 - 2r_P R t}} = \frac{1}{r_P\sqrt{1 + s^2 - 2st}}, \tag{2.37a}$$

where $s = R/r_P$ and $t = \cos\psi$; ψ being the centric angle between the radius vectors for $P_0(R, \varphi_0, \lambda_0)$ and $P(r_P, \varphi, \lambda)$ (see Fig. 2.4). The angle is related to the latitudes and longitudes of the points by the cosine formula from spherical trigonometry:

$$\cos\psi = \sin\varphi_0 \sin\varphi + \cos\varphi_0 \cos\varphi \cos(\lambda - \lambda_0). \tag{2.37b}$$

Equation (2.37a) can be expanded as a Taylor/binomial series in s, which can be re-arranged into a series in Legendre's polynomials $P_n(t)$ as:

$$l_{P0}^{-1} = \frac{1}{r_P}\sum_{n=0}^{\infty} s^n P_n(t); \quad s < 1. \tag{2.38a}$$

If P is located inside the sphere, the series becomes:

$$l_{P0}^{-1} = \frac{1}{R}\sum_{n=0}^{\infty} s^{-n} P_n(t); \quad s > 1. \tag{2.38b}$$

On the sphere (with s = 1), the series converges for $\psi \neq 0$:

$$l_{P0}^{-1} = \frac{1}{R}\sum_{n=0}^{\infty} P_n(t). \tag{2.38c}$$

Differentiate each side of Eq. (2.38a) w.r.t. r_P, multiply by $-2r_P$ and subtract the inverse distance. Then, after a few manipulations, one obtains *Poisson's kernel function*

$$\frac{r_P\left(r_P^2 - R^2\right)}{l_{P0}^3} = \sum_{n=0}^{\infty} \frac{(2n+1)R^n}{r_P^n} P_n(t); \quad r_P > R. \tag{2.39}$$

This function is important in solving Dirichlet's problem, the first boundary value problem of physical geodesy (see Sect. 2.7.5).

Exercise 2.1 Make a Taylor expansion of the inverse distance and compare with (2.38a) to show that $P_0(t) = 1$, $P_1(t) = t$ and $P_2(t) = (3t^2 - 1)/2$.

The solution is given in Appendix.

Legendre's polynomials have the following important properties:

$$|P_n| \le 1, \quad P_n(1) = 1, \quad P_n(-1) = (-1)^n \tag{2.40}$$

and

$$P_n(t) = \frac{2n-1}{n} t P_{n-1}(t) - \frac{n-1}{n} P_{n-2}(t); \quad n \ge 2. \tag{2.41}$$

By the recursive formula Eq. (2.41) the Legendre's polynomial can be determined numerically to any degree.

Legendre's polynomials are orthogonal in the interval −1 to 1, i.e.:

$$\int_{-1}^{1} P_n^2(t)dt = \frac{2}{2n+1} \quad \text{and} \quad \int_{-1}^{1} P_n(t)P_m(t)dt = 0, \quad \text{if } n \ne m. \tag{2.42}$$

Exercise 2.2 Verify (2.42) for n = 0–2.

The solution is given in Appendix.

Legendre's polynomials can also be determined by *Rodrigues's formula*:

$$P_n(t) = \frac{d^n(t^2-1)^n}{2^n n! dt^n}, \tag{2.43}$$

but this formula is less practical on a computer than the recursive formula.

From the ordinary Legendre's polynomials, the (*associated*) *Legendre functions* can be defined:

$$P_{nm}(t) = \left(1 - t^2\right)^{m/2} \frac{d^m P_n(t)}{dt^m}; \quad m \le n, \tag{2.44}$$

where m is the *order*. It is a basic component in spherical harmonics.

2.5 Spherical Harmonics

A fully normalized (surface) spherical harmonic of degree n and order m can be written (Sjöberg 1975, 1978):

$$Y_{nm}(\theta,\lambda) = N_{nm}P_{n|m|}(\cos\theta)\begin{cases}\sin m\lambda & if \quad m>0\\ \cos m\lambda & otherwise\end{cases}, \tag{2.45a}$$

where:

$$N_{nm} = \begin{cases}\sqrt{2(2n+1)\frac{(n-|m|)!}{(n+|m|)!}} & if \quad m\neq 0\\ \sqrt{2n+1} & m=0\end{cases} \tag{2.45b}$$

is a normalizing factor that makes the harmonics *orthonormal*, i.e., they obey:

$$\frac{1}{4\pi}\iint_{\sigma} Y_{nm}^2 d\sigma = 1, \tag{2.46}$$

where (θ,λ) are the co-latitude and longitude, and they are orthogonal to each other when averaged over the unit sphere (σ):

$$\frac{1}{4\pi}\iint_{\sigma} Y_{nm}Y_{kl}d\sigma = 0, \quad \text{if} \quad n\neq k \text{ and/or } m\neq l. \tag{2.47}$$

Note that we use the notations $Y_{nm}(\theta,\lambda), Y_{nm}, Y_{nm}(P)$ interchangeably when there could be no misunderstanding.

Let a distance between two points in space be given by $l_P = \sqrt{r_P^2 + r^2 - 2r_P rt}$, where $t = \cos\psi$ and r_P, r are the geocentric radii of the two points, separated by the geocentric angle ψ. Then Eqs. (2.38a, b) can be generalized to:

$$1/l_P = \sum_{n=0}^{\infty}\frac{r^n}{r_P^{n+1}}P_n(t); \quad r_P > r \quad \text{(external type series)} \tag{2.48a}$$

and

$$1/l_P = \sum_{n=0}^{\infty}\frac{r_P^n}{r^{n+1}}P_n(t), \quad \text{if} \quad r_P<r \quad \text{(internal type series)} \tag{2.48b}$$

- The *Legendre's polynomial* $P_n(\cos\psi)$ is related with the spherical harmonics by the *addition theorem*

$$(2n+1)P_n(\cos\psi) = \sum_{m=-n}^{n} Y_{nm}(P)Y_{nm}(Q), \tag{2.49}$$

where P and Q are the endpoints of an arc on the unit sphere of geocentric angle ψ.

- Any (decent) function f on a sphere can be expanded as a harmonic series:

$$f = f(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\theta, \lambda), \tag{2.50a}$$

where:

$$f_{nm} = \frac{1}{4\pi} \iint_{\sigma} f Y_{nm} d\sigma. \tag{2.50b}$$

- The function f may also be written as a *Laplace series*

$$f = f(\theta, \lambda) = \sum_{n=0}^{\infty} f_n(\theta, \lambda), \tag{2.51a}$$

where:

$$f_n(\theta, \lambda) = \frac{2n+1}{4\pi} \iint_{\sigma} f P_n(\cos\psi) d\sigma, \tag{2.51b}$$

or, using the addition theorem (2.49)

$$f_n(\theta, \lambda) = \sum_{m=-n}^{n} f_{nm} Y_{nm}(\theta, \lambda), \tag{2.51c}$$

which relates the Laplace harmonics to the spherical harmonics.

- The truncated series

$$\hat{f}(\theta, \lambda) = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} f_{nm} Y_{nm}(\theta, \lambda) \tag{2.52a}$$

has $1 + 3 + 5 + \cdots + (2n_{\max} + 1) = (1 + n_{\max})^2$ terms and an approximate resolution of:

$$v^{\circ} \approx 180^{\circ}/n_{\max}. \tag{2.52b}$$

For $n_{max} = 2159$ (which is the case for EGM2008; see below) there are 4.665.600 terms and a resolution of about 5′.

- The Newton integral of the Earth's potential with μ = gravitational constant times density is:

$$V(P) = \iint_{\sigma} \int_{0}^{r_s} \frac{\mu r^2}{l_P} dr d\sigma. \tag{2.53}$$

- It can be expanded in the external type harmonic series

$$V(P) = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} V_n(P), \quad r_P \geq (r_s)_{\max}, \tag{2.54a}$$

where:

$$V_n(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{0}^{r_s} \mu \frac{r^{n+2}}{R^{n+1}} dr Y_{nm} d\sigma. \tag{2.54b}$$

Here R is a selected radius (e.g. mean sea-level radius), and $r_s = r_s(\theta, \lambda)$ is the radius of the Earth's topography.

- If $r_P < (r_s)_{\max}$, the potential can be expanded in a combination of external and internal type series

$$V(P) = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} V_n^e(P) + \sum_{n=0}^{\infty} \left(\frac{r_P}{R}\right)^{n} V_n^i(P), \tag{2.55a}$$

where:

$$V_n^e(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{r_P}^{r_s} \mu \frac{r^{n+2}}{R^{n+1}} dr Y_{nm} d\sigma, \tag{2.55b}$$

and

$$V_n^i(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{r_P}^{r_s} \mu \frac{R^n}{r^{n-1}} dr Y_{nm} d\sigma. \tag{2.55c}$$

Note that the coefficients in (2.55b, 2.55c) change for each radius r_P.

Disregarding the atmosphere, the external type series in Eq. (2.55a) definitely converges outside the bounding sphere, the Brillouin sphere, which touches the peak of Mt. Chimborazo in Ecuador at elevation 6267 m with an Earth centre radius of 6384 km. Although, in the strict sense, the series is likely to diverge inside this sphere but be asymptotically divergent (Moritz 1980, Chaps. 6 and 7), it can be applied to very high degrees without notable commission errors also inside the sphere. However, when applied (analytically continued) inside topographic masses,

the series is biased and should be corrected for this error (see Sect. 5.2.6). This is in agreement with the approximation theorems of Runge-Krarup (Krarup 1969) and Keldysh-Laurentiev (Bjerhammar 1975), which states that there exist harmonic series that approximate the external type harmonic series arbitrarily well above the surface and down onto the surface of the Earth, respectively, and converge all the way down to the internal (Bjerhammar) sphere.

Integral formulas and equations on the sphere can frequently be derived and solved, respectively, in the spectral domain by spherical harmonics. For examples, see Sects. 3.3, and 8.2.

2.5.1 Spectral Filtering and Combination

2.5.1.1 Introduction

Here we derive the local least squares spectral filter for a stochastic function on the sphere based on the spectral representation of the observable and its error covariance matrix. Second, the local least squares spectral combination of two erroneous harmonic series is derived based on their full covariance matrices. In both problems, the transition from spectral representation of an estimator to an integral representation is demonstrated. Practical examples are given for the spectral filter and for the combination of a series and an integral formula.

Taking advantage of the full covariance matrices in the spectral combination implies a huge computational burden in determining the least squares filters and combinations for high degree spherical harmonic series. A reasonable compromise could be to consider only one weight parameter/degree, yielding the optimum filtering and combination of Laplace series as outlined in Sect. 7.5.

Spectral combination of harmonic functions has proved to be a practical tool to match various observables in physical geodesy (see Sects. 2.8.4, 4.4.4–4.4.6 and Chap. 6). Early models along this line were presented by Sjöberg (1979, 1980, 1981) as well as by Wenzel (1981). In Sjöberg (1979) and partly in Sjöberg (1980), integral formulas were presented for least squares combination of stochastic, random heterogeneous data, while otherwise, more realistically, only the errors of the data were considered stochastic. All these models have in common that the correlations between different spectral degrees of errors are disregarded, and frequently the models are based on minimizing the global variance or mean square error. See also Sjöberg (1984a, b, 1986), which provide the basics of least squares modification of Stokes' formula. Also, Sjöberg (2005) presents a local modification of Stokes' formula using weighting by degrees. Considering that Earth gravitational models (EGMs) are usually provided together with their full covariance matrices, at least up to some specific degrees, and that the qualities of the models vary over the surface of the Earth, all the information contained in the covariance matrices should be utilized in the combined solutions. This article derives such solutions for filtering and combination of EGMs, as well as in the combination of an EGM with an

integral formula. See also Sjöberg (2011a). Most of the following subsection is a reprint of Sjöberg (2011b).

2.5.1.2 Local Spectral Filtering

Let the gravity field related function v be developed into a finite series of spherical harmonics $Y_{nm}(\theta, \lambda)$ on the sphere

$$v = v(\theta, \lambda) = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} v_{nm} Y_{nm}(\theta, \lambda), \tag{2.56}$$

where v_{nm} are the harmonic coefficients, $n_{\max}$ is the maximum degree of expansion of the series and (θ, λ) is the (co-latitude, longitude) of the function. Consider the unbiased and biased estimators of v:

$$\tilde{v}' = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} \tilde{v}_{nm} Y_{nm}(\theta, \lambda) = \mathbf{e}^{\mathbf{T}} \mathbf{D} \tilde{\mathbf{v}}, \tag{2.57a}$$

and

$$\tilde{v} = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} p_{nm} \tilde{v}_{nm} Y_{nm}(\theta, \lambda) = \mathbf{p}^{\mathbf{T}} \mathbf{D} \tilde{\mathbf{v}} \tag{2.57b}$$

where p_{nm} is weighting parameter to be fixed,

$$\tilde{v}_{nm} = v_{nm} + dv_{nm} \quad \text{with} \quad E\{dv_{nm}\} = 0, \tag{2.57c}$$

and $\mathbf{D}$ is a diagonal matrix with $(n_{\max} + 1)^2$ elements $(Y_{00}, Y_{1-1}, \ldots, Y_{n_{\max} n_{\max}})$. Furthermore $\mathbf{p}$ and $\tilde{\mathbf{v}}$ are vectors with obvious elements, and $\mathbf{e}^{\mathbf{T}} = (1, 1, \ldots, 1)$.

The error of the unbiased and biased estimators can be written:

$$d\tilde{v}' = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} dv_{nm} Y_{nm}(\theta, \lambda) = \mathbf{e}^{\mathbf{T}} \mathbf{D} \mathbf{dv} \tag{2.58a}$$

and

$$d\tilde{v} = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} [p_{nm} dv_{nm} + v_{nm}(p_{nm} - 1)] Y_{nm}(\theta, \lambda) = \mathbf{p}^{\mathrm{T}} \mathbf{D} \mathbf{dv} + \left(\mathbf{p}^{\mathbf{T}} - \mathbf{e}^{\mathbf{T}}\right) \mathbf{D} \mathbf{v}. \tag{2.58b}$$

Assuming that the covariance matrix of the error vector **dv** is **Q**, one obtains the following variance and Mean Square Error (MSE) of $\tilde{v}'$ and $\tilde{v}$, respectively:

$$\sigma_{v'}^2 = \mathbf{e^T DQDe} \tag{2.59a}$$

and

$$\mathrm{MSE}(\tilde{v}) = \mathbf{p^T DQDp} + \left(\mathbf{p^T} - \mathbf{e^T}\right)\mathbf{Dvv^T D}(\mathbf{p} - \mathbf{e}). \tag{2.59b}$$

The spectral filter solutions are provided by Eq. (2.2b), and the optimum filter is the one where the filter parameters p_{nm} are chosen such that the MSE is a minimum. This choice for filter parameters is thus obtained by differentiating the MSE w.r.t. **p** and equating it to zero. The result is the matrix equation

$$\left(\mathbf{DQD} + \mathbf{Dvv^T D}\right)\mathbf{p} - \mathbf{Dvv^T De} = \mathbf{0} \tag{2.60}$$

with the solution

$$\hat{\mathbf{p}} = \left(\mathbf{DQD} + \mathbf{Dvv^T D}\right)^{-1}\mathbf{Dvv^T De}. \tag{2.61}$$

Hence, by inserting Eq. (2.61) into Eq. (2.2b), the Local Least Squares Spectral Filter becomes:

$$\hat{v} = \mathbf{p^T D\tilde{v}} = \mathbf{e^T}\left[\mathbf{I} - \mathbf{DQD}\left(\mathbf{DQD} + \mathbf{Dvv^T D}\right)^{-1}\right]\mathbf{D\tilde{v}} \tag{2.62a}$$

with the mean square error

$$\mathrm{MSE}\{\hat{v}\} = \mathbf{e^T DQDe} - \mathbf{e^T DQD}\left(\mathbf{DQD} + \mathbf{Dvv^T D}\right)^{-1}\mathbf{DQDe}, \tag{2.62b}$$

and Eqs. (2.62a, b) can also be simplified to:

$$\hat{v} = \mathbf{e^T D\tilde{v}} - \mathbf{e^T DQ}\left(\mathbf{Q} + \mathbf{vv^T}\right)^{-1}\tilde{\mathbf{v}} \tag{2.63a}$$

and

$$\mathrm{MSE}\{\hat{v}\} = \mathbf{e^T DQDe} - \mathbf{e^T DQ}\left(\mathbf{Q} + \mathbf{vv^T}\right)^{-1}\mathbf{QDe}. \tag{2.63b}$$

Equations (2.59a) and (2.63b) show that the MSE of $\hat{v}$ is smaller than the variance of v'. As $\mathbf{vv^T}$ is not known, there is a practical problem in applying Eqs. (2.63a, b). However, $\mathbf{Q} + \mathbf{vv^T}$ is unbiasedly estimated by $\tilde{\mathbf{v}}\tilde{\mathbf{v}}^\mathbf{T}$, and, by this substitution, the filter and its covariance matrix can be realized.

2.5.1.3 Generalized Filtering

Here we assume that a function $w(\theta,\lambda)$ is related to function $v(\theta,\lambda)$ by the harmonic series

$$w = w(\theta,\lambda) = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} k_{nm} v_{nm} Y_{nm}(\theta,\lambda), \tag{2.64}$$

where k_{nm} are known coefficients (possibly functions of radial position). In analogy with above, the general estimator

$$\tilde{w} = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} q_{nm} \tilde{v}_{nm} Y_{nm}(\theta,\lambda) = \mathbf{q}^{\mathbf{T}}\mathbf{D}\tilde{\mathbf{v}} \tag{2.65}$$

is optimized in the least squares sense by the weight vector

$$\hat{\mathbf{q}} = \left(\mathbf{DQD} + \mathbf{Dvv^{T}D}\right)^{-1}\mathbf{Dvv^{T}Dk} = \left[\mathbf{I} - \mathbf{DQD}\left(\mathbf{DQD} + \mathbf{Dvv^{T}D}\right)^{-1}\right]\mathbf{k}, \tag{2.66}$$

yielding the least squares estimator

$$\hat{w} = \mathbf{k^{T}D}\tilde{\mathbf{v}} - \mathbf{k^{T}DQ}\left(\mathbf{Q} + \mathbf{vv^{T}}\right)^{-1}\tilde{\mathbf{v}}, \tag{2.67a}$$

with the minimum mean square error (with respect to choice of $\mathbf{q}$)

$$\mathrm{MSE}\{\hat{w}\} = \mathbf{k^{T}DQDk} - \mathbf{k^{T}DQ}\left(\mathbf{Q} + \mathbf{vv^{T}}\right)^{-1}\mathbf{QDk}. \tag{2.67b}$$

Comparing Eqs. (2.64) and (2.67a), we notice that $\hat{q}_{nm} = \hat{p}_{nm}$, which we will take advantage of in the integral representations that follow below.

2.5.1.4 Integral Representation of the Filter

Assuming that the spherical harmonics $Y_{nm}(\theta,\lambda)$ are fully normalized, it means that they are mutually orthonormal, i.e.:

$$\iint_{\sigma} Y_{nm} Y_{kl} d\sigma = \begin{cases} 4\pi, & if \quad (n,m) = (k,l) \\ 0 & otherwise. \end{cases} \tag{2.68}$$

Then we can express Eq. (2.67a) by the integral

$$\hat{w}_P = \frac{1}{4\pi} \iint_{\sigma} K(P,Q)\tilde{v} d\sigma, \tag{2.69a}$$

where the kernel function $K(P,Q)$ becomes:

$$K(P,Q)=\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}k_{nm}\hat{p}_{nm}Y_{nm}(\theta_P,\lambda_P)Y_{nm}(\theta_Q,\lambda_Q). \tag{2.69b}$$

Here Q is the integration point on the unit sphere (denoted by σ).

Example 2.1 Use Eq. (2.69a) to estimate the disturbing potential T_P at the radius r_P from the gravity anomaly $\Delta\tilde{g}$ on the sphere of radius R.

Solution Choosing $k_{nm}=0$ for $n<2$ and $k_{nm}=r_P/(n-1)(R/r_P)^{n+1}$ for $n\geq 2$, the solution is obtained by:

$$\hat{T}_P=\frac{r_P}{4\pi}\iint_{\sigma}S(P,Q)\Delta\tilde{g}d\sigma, \tag{2.70a}$$

where:

$$S(P,Q)=\sum_{n=2}^{n_{max}}\frac{1}{n-1}\left(\frac{R}{r_P}\right)^{n+1}\sum_{m=-n}^{n}\hat{p}_{nm}Y_{nm}(P)Y_{nm}(Q). \tag{2.70b}$$

2.5.1.5 Local Spectral Combination

Let $\tilde{u}$ and $\tilde{v}$ be two unbiased estimators of the finite harmonic series v, given by Eq. (2.56). Let the estimators be expressed by the series

$$\tilde{u}=\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}\tilde{u}_{nm}Y_{nm}(\theta,\lambda)=\mathbf{e}^{\mathbf{T}}\mathbf{D}\tilde{\mathbf{u}} \tag{2.71a}$$

and

$$\tilde{v}=\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}\tilde{v}_{nm}Y_{nm}(\theta,\lambda)=\mathbf{e}^{\mathbf{T}}\mathbf{D}\tilde{\mathbf{v}}, \tag{2.71b}$$

where both sets of coefficients $\tilde{u}_{nm}$ and $\tilde{v}_{nm}$ are unbiased estimates of v_{nm} with random errors du_{nm} and dv_{nm}, respectively, and the last parts of the equations are obvious matrix representations. We will assume also that the errors of the coefficients have expectations zero, and their covariance matrices will be denoted:

$$E\{\mathbf{dudu^T}\} = \mathbf{\Sigma}, \quad E\{\mathbf{dvdv^T}\} = \mathbf{Q}, \quad \text{and} \quad E\{\mathbf{dudv^T}\} = \mathbf{\Omega}. \tag{2.72}$$

The general unbiased spectral combination of the two series of Eqs. (2.71a, b) can be written:

$$\tilde{w} = \mathbf{p^T D\tilde{u}} + (\mathbf{e^T} - \mathbf{p^T})\mathbf{D\tilde{v}}, \tag{2.73}$$

where, again, $\mathbf{p}$ is a weight vector. The error and variance of this estimator become:

$$dw = \mathbf{p^T Ddu} + (\mathbf{e^T} - \mathbf{p^T})\mathbf{Ddv} \tag{2.74}$$

and

$$\begin{aligned}\sigma_w^2 = E\{dw^2\} = \mathbf{p^T D\Sigma Dp} + (\mathbf{e^T} - \mathbf{p}^T)\mathbf{DQD}(\mathbf{e} - \mathbf{p}) \\ + \mathbf{p^T D\Omega D}(\mathbf{e} - \mathbf{p}) + (\mathbf{e^T} - \mathbf{p^T})\mathbf{D\Omega^T Dp}.\end{aligned} \tag{2.75}$$

The least squares choice of $\mathbf{p}$ minimizes the variance, and this minimum is attained by differentiating Eq. (2.75) w.r.t. $\mathbf{p}$ and equating to zero. The result is:

$$\mathbf{DMDp} - \mathbf{D}(\mathbf{Q} - \mathbf{\Omega})\mathbf{De} = \mathbf{0}, \tag{2.76a}$$

where:

$$\mathbf{M} = \mathbf{\Sigma} + \mathbf{Q} - \mathbf{\Omega} - \mathbf{\Omega^T}, \tag{2.76b}$$

with the solution

$$\hat{\mathbf{p}} = (\mathbf{DMD})^{-1}\mathbf{D}(\mathbf{Q} - \mathbf{\Omega})\mathbf{De} = \mathbf{D}^{-1}\mathbf{M}^{-1}(\mathbf{Q} - \mathbf{\Omega})\mathbf{De}. \tag{2.76c}$$

Hence, the least squares spectral combination becomes:

$$\hat{w} = \mathbf{e^T D}(\mathbf{Q} - \mathbf{\Omega^T})\mathbf{M}^{-1}\tilde{\mathbf{u}} + \mathbf{e^T D}(\mathbf{\Sigma} - \mathbf{\Omega})\mathbf{M}^{-1}\tilde{\mathbf{v}} \tag{2.77a}$$

with the variance

$$\sigma_{\hat{w}}^2 = \mathbf{e^T DQDe} - \mathbf{e^T D}(\mathbf{Q} - \mathbf{\Omega^T})\mathbf{M}^{-1}(\mathbf{Q} - \mathbf{\Omega})\mathbf{De}. \tag{2.77b}$$

2.5.1.6 Generalization

Let us assume that the function

$$w = w(r,\theta,\lambda) = \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} w_{nm} Y_{nm}(\theta,\lambda), \tag{2.78}$$

where $r \geq R$, is a 3D function, unbiasedly estimable on and outside the sphere of radius R by the functions $\tilde{u}(\theta,\lambda)$ and $\tilde{v}(\theta,\lambda)$ on the sphere. Then the general unbiased estimators for w can be written:

$$\tilde{w}_1 = \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} f_{nm} \tilde{u}_{nm} Y_{nm}(\theta,\lambda) = \mathbf{f^T D\tilde{u}} \tag{2.79a}$$

and

$$\tilde{w}_2 = \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} g_{nm} \tilde{v}_{nm} Y_{nm}(\theta,\lambda) = \mathbf{g^T D\tilde{v}}, \tag{2.79b}$$

where the given coefficients f_{nm} and g_{nm} bring the harmonics of u and v to those of w, i.e. $w_{nm} = f_{nm} u_{nm} = g_{nm} v_{nm}$. Also, vectors $\mathbf{f}$ and $\mathbf{g}$ have the elements $(R/r)^{n+1} f_{nm}$ and $(R/r)^{n+1} g_{nm}$, with $0 \leq n \leq n_{\max}$ and $-n \leq m \leq n$.

A general unbiased combined estimator for w can be written:

$$\tilde{w} = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} p_{nm} f_{nm} \tilde{u}_{nm} Y_{nm}(\theta,\lambda) + \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} (1 - p_{nm}) g_{nm} \tilde{v}_{nm} Y_{nm}(\theta,\lambda), \tag{2.80a}$$

where p_{nm} are arbitrary degree/order weights. With matrix notations the estimator becomes:

$$\tilde{w} = \mathbf{d}^T \tilde{\mathbf{u}} + \left(\mathbf{g}^T \mathbf{D} - \mathbf{h}^T\right) \tilde{\mathbf{v}}, \tag{2.80b}$$

where $\mathbf{d} = \mathbf{Fp}$, $\mathbf{h} = \mathbf{Gp}$ and $\mathbf{p}$ is the vector with elements p_{nm}. Here $\mathbf{F}$ and $\mathbf{G}$ are diagonal matrices with elements from vectors $\mathbf{Df}$ and $\mathbf{Dg}$, respectively.

The variance of $\tilde{w}$ becomes:

$$\sigma_{\tilde{w}}^2 = \mathbf{d}^T \mathbf{\Sigma d} + \left(\mathbf{g}^T \mathbf{D} - \mathbf{h}^T\right) \mathbf{Q} (\mathbf{Dg} - \mathbf{h}) + \mathbf{d}^T \mathbf{\Omega} (\mathbf{Dg} - \mathbf{h}) + \left(\mathbf{g}^T \mathbf{D} - \mathbf{h}^T\right) \mathbf{\Omega}^T \mathbf{d}^T, \tag{2.81}$$

and its minimum is obtained by differentiating Eq. (2.81) w.r.t. $\mathbf{p}$ and equating it to zero. The result is:

$$\mathbf{Hp} - \mathbf{GQDg} + \mathbf{F\Omega Dg} = \mathbf{0}, \tag{2.82a}$$

where:

$$\mathbf{H} = \mathbf{F\Sigma F} + \mathbf{GQG} - \mathbf{F\Omega G} - \mathbf{G\Omega}^T\mathbf{F}. \tag{2.82b}$$

Hence, the optimum weight vector becomes:

$$\hat{\mathbf{p}} = \mathbf{H}^{-1}(\mathbf{GQ} - \mathbf{F\Omega})\mathbf{Dg}, \tag{2.83}$$

yielding the optimum estimator for w

$$\hat{w} = \hat{\mathbf{p}}^T\mathbf{F}\tilde{\mathbf{u}} + \left(\mathbf{g}^T\mathbf{D} - \hat{\mathbf{p}}^T\mathbf{G}\right)\tilde{\mathbf{v}}, \tag{2.84a}$$

with the variance

$$\sigma_{\hat{w}}^2 = \mathbf{g}^T\mathbf{DQDg} - \hat{\mathbf{p}}^T\mathbf{H}\hat{\mathbf{p}} = \mathbf{g}^T\mathbf{D}\left[\mathbf{Q} - \left(\mathbf{QG} - \mathbf{\Omega}^T\mathbf{F}\right)\mathbf{H}^{-1}(\mathbf{GQ} - \mathbf{F\Omega})\right]\mathbf{Dg}. \tag{2.84b}$$

2.5.1.7 Integral and Series Combination

The estimator $\tilde{w}_2$ of Eq. (2.79b) can be expressed by the integral

$$\tilde{w}_2 = \frac{1}{4\pi}\iint_{\sigma} \widetilde{M}(r, \theta, \lambda, \theta', \lambda')\tilde{v}(\theta', \lambda')d\sigma', \tag{2.85a}$$

where the kernel function is given by:

$$\widetilde{M}(r, \theta, \lambda, \theta', \lambda') = \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} g_{nm} Y_{nm}(\theta, \lambda) Y_{nm}(\theta', \lambda'). \tag{2.85b}$$

It follows that Eq. (2.80a) can be rewritten as

$$\hat{w} = \frac{1}{4\pi}\iint_{\sigma} M(r, \theta, \lambda, \theta', \lambda')\tilde{v}(\theta', \lambda')d\sigma + \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} \hat{p}_{nm} f_{nm} \tilde{u}_{nm} Y_{nm}(\theta, \lambda), \tag{2.86a}$$

where:

$$M(r, \theta, \lambda, \theta', \lambda') = \sum_{n=0}^{n_{\max}} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} (1 - \hat{p}_{nm}) g_{nm} Y_{nm}(\theta, \lambda) Y_{nm}(\theta', \lambda'). \tag{2.86b}$$

Example 3.1 Geoid height estimation from an EGM and an integral formula with estimated gravity anomaly $\Delta\tilde{g}$.

The least squares estimator for the geoidal height is given by Eqs. (2.86a, b) for $f_{nm} = R/\gamma$, $\tilde{u}_{nm} = \tilde{T}_{nm}$, $\tilde{v} = \Delta\tilde{g}$, $g_{nm} = R/(n-1)$, $r = R$, and the degree summation in Eq. (2.86b) starts at $n_{\min} = 2$. The result is:

$$\widehat{N} = \frac{R}{4\pi\gamma}\iint_{\sigma} S(\theta,\lambda,\theta',\lambda')\Delta\tilde{g}d\sigma + \frac{R}{\gamma}\sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n}\hat{p}_{nm}\widetilde{T}_{nm}Y_{nm}(\theta,\lambda), \tag{2.87a}$$

where:

$$S(\theta,\lambda,\theta',\lambda') = \sum_{n=2}^{n_{\max}}\frac{1}{n-1}\sum_{m=-n}^{n}(1-\hat{p}_{nm})Y_{nm}(\theta,\lambda)Y_{nm}(\theta',\lambda'). \tag{2.87b}$$

The least squares weights are given by Eq. (2.83) when considering the above choices of f_{nm} and g_{nm}.

2.5.1.8 Filtering and Weighting by Laplace Harmonics

So far we considered filtering and weighting by spherical harmonics. From a numerical point of view, when considering the large dimension of the matrices to be inverted, e.g. $\mathbf{Q} + \mathbf{v}\mathbf{v}^T$ in Eq. (2.63a), it could be reasonable to modify the technique to one weight factor/degree. This is obtained by considering that Eq. (2.56) can be written as the series of Laplace harmonics

$$x_n = x_n(\theta,\lambda) = \sum_{m=-n}^{n} v_{nm}Y_{nm}(\theta,\lambda), \tag{2.88}$$

which yields:

$$v = \sum_{n=0}^{n_{\max}} x_n. \tag{2.89}$$

Similarly, the estimator of Eq. (2.57b) and its error can be expressed as:

$$\tilde{v} = \sum_{n=0}^{n_{\max}} p_n\tilde{x}_n = \mathbf{p}^T\tilde{\mathbf{x}} \tag{2.90a}$$

and

$$d\tilde{v} = \sum_{n=0}^{n_{\max}} [p_n d\tilde{x}_n + (p_n - 1)x_n] = \mathbf{p}^T \mathbf{dx} + \left(\mathbf{p}^T - \mathbf{e}^T\right)\mathbf{x}. \tag{2.90b}$$

Hence, the MSE of $\tilde{v}$ becomes:

$$\mathrm{MSE}(\tilde{v}) = \mathbf{p^T Q_{xx} p} + \left(\mathbf{p^T} - \mathbf{e^T}\right)\mathbf{x}\mathbf{x}^T(\mathbf{p} - \mathbf{e}) \tag{2.91}$$

with the least squares choice for the weight vector:

$$\hat{\mathbf{p}} = \left(\mathbf{Q_{xx}} + \mathbf{x}\mathbf{x}^T\right)^{-1}\mathbf{x}\mathbf{x}^T\mathbf{e}. \tag{2.92}$$

Finally, the filtered estimator becomes:

$$\hat{v} = \mathbf{e^T}\left[\mathbf{I} - \mathbf{Q_{xx}}\left(\mathbf{Q_{xx}} + \mathbf{x}\mathbf{x^T}\right)^{-1}\right]\tilde{\mathbf{x}} \tag{2.93a}$$

with the MSE

$$\mathrm{MSE}\{\hat{v}\} = \mathbf{e^T Q_{xx} e} - \mathbf{e^T Q_{xx}}\left(\mathbf{Q_{xx}} + \mathbf{v}\mathbf{v^T}\right)^{-1}\mathbf{Q_{xx} e}. \tag{2.93b}$$

Similarly the generalized filter of Eq. (2.64) can be obtained for the restriction of the number of weights to one/degree, and the integral representation of the filter becomes:

$$\hat{w}_P = \frac{1}{4\pi}\iint_{\sigma} K(P, Q)\tilde{x}d\sigma, \tag{2.94a}$$

where:

$$K(P, Q) = \sum_{n=0}^{n_{\max}} (2n + 1)k_n\hat{p}_n P_n(\cos\psi). \tag{2.94b}$$

Here $P_n(\cos\psi)$ is the n-th Legendre's polynomial, and ψ is the geocentric angle between the computation and integration points.

Finally the least square spectral combination for degree weighting corresponding to Eqs. (2.77a, b) can be written:

$$\hat{w} = \mathbf{e^T}\left(\mathbf{Q} - \mathbf{\Omega^T}\right)\mathbf{M}^{-1}\tilde{\mathbf{u}} + \mathbf{e^T}(\mathbf{\Sigma} - \mathbf{\Omega})\mathbf{M}^{-1}\tilde{\mathbf{v}} \tag{2.95a}$$

with the variance:

$$\sigma_{\tilde{w}}^2 = \mathbf{e}^{\mathbf{T}}\mathbf{Q}\mathbf{e} - \mathbf{e}^{\mathbf{T}}\left(\mathbf{Q}-\mathbf{\Omega}^{\mathbf{T}}\right)\mathbf{M}^{-1}(\mathbf{Q}-\mathbf{\Omega})\mathbf{e}, \tag{2.95b}$$

where $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ now means vectors of Laplace harmonics, and $\mathbf{Q}, \mathbf{\Sigma}, \mathbf{\Omega}$ are the corresponding covariance matrices, and $\mathbf{M}$ is again defined by Eq. (2.76b).

2.5.1.9 Conclusions

The above solutions are the locally optimum spectral filters and combinations of functions on the sphere in the sense of minimum MSE. They utilize the full covariance matrices of the stochastic errors of the parameters representing the functions. In the most advanced cases, this implies that the total covariance matrix of a given EGM is employed, implying a considerable computational burden. This workload can be relaxed by considering only spectral weighting by degree, yielding the filter and spectral combination of Laplace series (see e.g. Sects. 2.8.4 and 7.6).

The above study includes the theoretical derivations of general filters and spectral combinations of harmonic series or a harmonic series and an integral formula, and the solutions should be suitable for solving both direct and inverse problems on the sphere.

2.6 Ellipsoidal Harmonics

As the Earth is rather a two-axis ellipsoid than a sphere, ellipsoidal harmonics are better suited than spherical ones for global modelling. The relation between Cartesian and ellipsoidal coordinates was presented in Eq. (2.22). The Laplace equation in ellipsoidal coordinates and its solution for the exterior case were derived in Heiskanen and Moritz (1967, Sects. 1–19 and 1–20). The solutions are:

$$\begin{aligned}\left(u^2 + E^2\cos^2\theta\right)\Delta V = {} & \left(u^2+E^2\right)\frac{\partial^2 V}{\partial u^2} + 2u\frac{\partial V}{\partial u} + \frac{\partial^2 V}{\partial \theta^2} \\ & + \cot\theta\frac{\partial V}{\partial \theta} + \frac{u^2+E^2\cos^2\theta}{\left(u^2+E^2\right)\sin^2\theta}\frac{\partial^2 V}{\partial \lambda^2} = 0\end{aligned} \tag{2.96a}$$

and

$$V_e(u,\theta,\lambda) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\frac{Q_{nm}(iu/E)}{Q_{nm}(ib/E)}\tilde{A}_{nm}Y_{nm}(\theta,\lambda), \tag{2.96b}$$

where $Q_{nm}(\,)$ are associate Legendre's polynomials of the 2nd kind, and $\tilde{A}_{nm}$ are normalized harmonic coefficients on the ellipsoid, while Y_{nm} are the surface harmonics. This implies that the potential on the ellipsoid can be expressed:

$$V_e(\mathrm{b},\theta,\lambda)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\widetilde{A}_{nm}Y_{nm}(\theta,\lambda), \tag{2.97}$$

where $\Omega=4\pi ab$ is the area of the ellipsoid with a and b being the semi-axes of the ellipsoid. Starting from Eq. (2.96b), the disturbing potential, the gravity anomaly and other gravity related quantities can be represented in ellipsoidal harmonics. Equation (2.97) indicates that a series of surface spherical harmonics can represent a function (not necessarily a potential) on a rather arbitrary surface.

One important application of ellipsoidal harmonics was utilized in the development of the Earth Gravitational Model 2008 (Pavlis et al. 2012). In this technique, which was already tested in earlier OSU EGMs, a preliminary EGM is first determined from a global set of 5′ equal-area mean gravity anomalies analytically downward extended to the reference ellipsoid. From this set of data, "terrestrial" ellipsoidal potential coefficients were solved from an overdetermined linear system of equations of coefficients $\tilde{A}_{nm}$ from harmonic degree 2 complete to degree and order 2159, and these coefficients were transformed to spherical harmonics by the method of Gleason (1988, Eq. 2.10). In a second least squares adjustment, the satellite only spherical harmonic model ITG-GRACE03S, complete to degree 180, and the "terrestrial" harmonics were merged into a final solution.

2.7 Fundamentals of Potential Theory

2.7.1 Basic Concepts and Formulas

Newton's law of gravitation is fundamental to potential theory. It states that (the magnitude of) the attracting force F between two point masses m_1 and m_2 at distance l is:

$$F=Gm_1m_2/l^2, \tag{2.98}$$

where G ($\approx 6.67\times10^{-11}$ $\mathrm{Nm^2/kg^2}$) is the gravitational constant. From now on, we set the attracted mass m_1 to 1 (unit mass) at point P (x, y, z) and the attracting mass $m_2=m$ at point Q (ξ,η,ζ). Then the 3-D attraction force at P can be expressed:

$$\mathbf{F}=-\mathrm{G}\frac{m}{l^3}\mathbf{r}, \tag{2.99}$$

or, in Cartesian components (see Fig. 2.5):

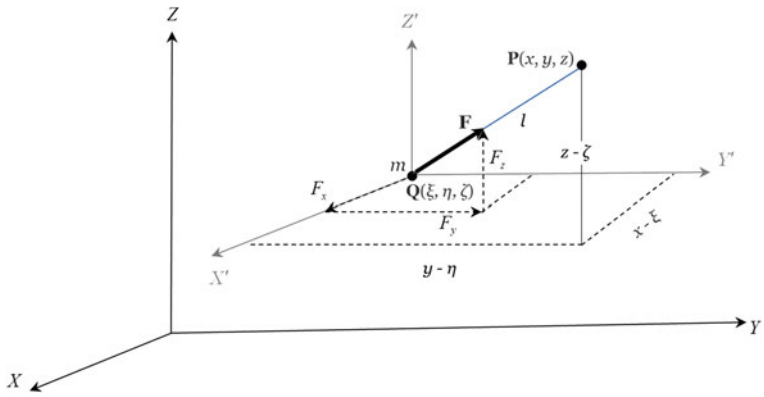

Fig. 2.5 Components of the attraction force $\mathbf{F} = \sqrt{F_x^2 + F_y^2 + F_z^2}$

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = -G\frac{m}{l^3}\begin{pmatrix} x-\xi \\ y-\eta \\ z-\zeta \end{pmatrix}. \tag{2.100}$$

The potential energy V that the unit mass experiences at point P is called *the gravitational potential* at P:

$$V = G\frac{m}{l}, \tag{2.101}$$

and the potential is thus related to the vector of attraction (gravitation) by the equation

$$\mathbf{F} = grad(V) = \begin{pmatrix} \frac{\partial V}{\partial x} & \frac{\partial V}{\partial y} & \frac{\partial V}{\partial y} \end{pmatrix}^T \tag{2.102}$$

and its magnitude becomes:

$$F = \sqrt{\left(\frac{\partial V}{\partial x}\right)^2 + \left(\frac{\partial V}{\partial y}\right)^2 + \left(\frac{\partial V}{\partial z}\right)^2}. \tag{2.103}$$

The potential is a scalar quantity, and it is additive. The latter property implies that the potential at a point P_j can be determined as the sum of the potentials generated by all surrounding point masses m_i:

$$V_j = G\sum_{i=1}^{n}\frac{m_i}{l_{ji}}, \tag{2.104a}$$

where l_{ji} is the distance between points P_j and P_i. This equation can be generalized to a closed body with volume υ and density ρ:

$$V_j = G \iiint_{\upsilon} \frac{dm}{l} = G \iiint_{\upsilon} \frac{\rho d\upsilon}{l}, \tag{2.104b}$$

where $l = l_{ji}$. According to Eq. (2.102), the force vector becomes:

$$\mathbf{F}_j = grad(V_j) = G \iiint_{\upsilon} \rho \left(\frac{\partial}{\partial x_j}, \quad \frac{\partial}{\partial y_j}, \quad \frac{\partial}{\partial z_j} \right)^T \frac{1}{l} d\upsilon. \tag{2.105}$$

- If the potential is generated by a materialized surface S with surface density $\kappa = dm/dS$, the correspondence to the 3D-Newton integral (2.104b) is given by

$$V = G \iint_{S} \frac{\kappa}{l} dS. \tag{2.106a}$$

  This potential is continuous on and outside the surface, but its derivatives are discontinuous at the surface. Hence, the normal derivative on the surface becomes

$$\frac{\partial V}{\partial n} = G \iint_{S} \kappa \frac{\partial}{\partial n} \left(\frac{1}{l} \right) dS \mp 2\pi G\kappa, \tag{2.106b}$$

  where the minus/plus sign applies to the exterior/internal side of the surface.

- Introducing $\mu =$ gravitational constant times topographic density into Eq. (2.104b), the Newton integral of the Earth's potential, can be written:

$$V(P) = \iint_{\sigma} \int_{0}^{r_s} \frac{\mu r^2}{l_P} dr d\sigma, \tag{2.107}$$

  where σ is the unit sphere and r_s is the radius of the Earth's surface.

- The potential can be expanded in the external-type harmonic series

$$V(P) = \sum_{n=0}^{\infty} \left(\frac{R}{r_P} \right)^{n+1} V_n(P), \quad r_P \geq (r_s)_{\max}, \tag{2.108a}$$

  where:

$$V_n(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{0}^{r_s} \mu \frac{r^{n+2}}{R^{n+1}} dr Y_{nm} d\sigma. \tag{2.108b}$$

Here R is a selected radius (e.g. mean sea level radius), and $r_s = r_s(\theta, \lambda)$.

- If $r_P < (r_s)_{\max}$, the potential can be expanded in a combination of external and internal type series:

$$V(P) = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} V_n^e(P) + \sum_{n=0}^{\infty} \left(\frac{r_P}{R}\right)^n V_n^i(P), \tag{2.109a}$$

where:

$$V_n^e(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{r_P}^{r_s} \mu \frac{r^{n+2}}{R^{n+1}} dr Y_{nm} d\sigma, \tag{2.109b}$$

and

$$V_n^i(P) = \sum_{m=-n}^{n} \frac{Y_{nm}(P)}{2n+1} \iint_{\sigma} \int_{r_P}^{r_s} \mu \frac{R^n}{r^{n-1}} dr Y_{nm} d\sigma. \tag{2.109c}$$

Note that the coefficients in (2.109b, 2.109c) change for each radius r_P.

2.7.2 *Laplace's and Poisson's Equations*

A function V is called harmonic if it satisfies Laplace's equation:

$$\Delta V = 0, \quad \text{where} \quad \Delta = grad \cdot grad = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{2.110}$$

Proposition 2.1 *Every gravitational potential is harmonic outside the attracting masses.*

Proof Applying the Laplace operator to Eq. (2.104b), one obtains:

$$\Delta V_P = \iiint_{\upsilon} \rho \Delta \left(\frac{1}{l_P}\right) d\upsilon = 0. \tag{2.111}$$

This is because as P is located outside the masses, $\Delta(1/l_P) = 0$ for all integration points in the body.

Corollary 2.1 (Poisson's differential equation)
If P is located inside the attracting masses, then $\Delta V_P = -4\pi G \rho_P$.

Proof The proof will be presented as an application of Gauss' theorem below.

2.7.3 Laplace's Equation and Its Solution in Spherical Coordinates

Laplace's equation in spherical coordinates reads (Heiskanen and Moritz 1967, p. 20):

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V}{\partial r}\right)+\frac{1}{r^2}\Delta^* V=0, \tag{2.112a}$$

where Δ^* is the *Beltrami operator* ("the Laplace operator on the sphere") defined by:

$$\Delta^*=\frac{\partial^2}{\partial\theta^2}+\cot\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\lambda^2}. \tag{2.112b}$$

By introducing V as a product of two functions that separate the variables (variable separation):

$$V(r,\theta,\lambda)=R(r)Y(\theta,\lambda), \tag{2.113}$$

Equation (2.112a) can be written:

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)=-\frac{\Delta^* Y}{Y}. \tag{2.114}$$

The two members can be separated by adding an arbitrary constant to each member. As will be shown, successful solutions are obtained by subtracting the constant $n(n+1)$ to both members of Eq. (2.114), which yields the two equations

$$\frac{1}{R}\frac{\partial}{\partial r}\left(r^2\frac{\partial R}{\partial r}\right)-n(n+1)R=0 \tag{2.115a}$$

and

$$\Delta^* Y+n(n+1)Y=0. \tag{2.115b}$$

It can easily be checked that each solution to (2.115a) is an arbitrary constant times r^n or $r^{-(n+1)}$. Setting also $Y=Y_n$ in (2.115b), one can show that its solution for each n is a set of solutions of the surface spherical harmonics $Y_{nm}(\theta,\lambda)$, where m ranges from $-n$ to plus n.

Adding all partial solutions to V, its general solution becomes the sum of the external and internal types of solutions in solid spherical harmonics:

$$V = V^e + V_i, \tag{2.116a}$$

where:

$$V^e = \sum_{n=0}^{\infty} r^{-(n+1)} \sum_{m=-n}^{n} A_{nm} Y_{nm}(\theta, \lambda) \tag{2.116b}$$

and

$$V^i = \sum_{n=0}^{\infty} r^n \sum_{m=-n}^{n} B_{nm} Y_{nm}(\theta, \lambda) \tag{2.116c}$$

are the external and internal solutions to Laplace's equation, respectively. Here the constants A_{nm} and B_{nm} are arbitrary, to be fixed for each specific problem with gravitational masses at hand. On or outside the Brillouin sphere (surrounding all masses), all B_{nm} are zero, while for the topographic potential representation at or below the geoid (approximated by a sphere), all A_{nm} vanish.

Note that Eq. (2.115b) implies that:

$$\Delta^* Y_{nm} = -n(n+1) Y_{nm}, \quad \text{for all} \quad n \text{ and } m. \tag{2.117}$$

2.7.4 Gauss' and Green's Integral Formulas

Gauss' and Green's formulas are basic formulas for potential theory. Here we present some of their varieties.

Gauss' (divergence) theorem for a closed volume υ with surface S applied to the vector $\mathbf{F}$ reads:

$$\iiint_{\upsilon} div\mathbf{F} d\upsilon = \iint_{S} \mathbf{F} \cdot \mathbf{n} dS, \tag{2.118}$$

where $\mathbf{n}$ is the external normal unit vector to the surface S and

$$div\mathbf{F} = \nabla \cdot \mathbf{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} \tag{2.119}$$

is *the divergence* of the vector $\mathbf{F}$.

Assuming as above that $\mathbf{F} = grad(V)$, then $\mathrm{div}\mathbf{F} = \Delta V$, and one obtains *Gauss' integral formula* for the potential as:

$$\iiint_{\upsilon} \mathrm{div}\mathbf{F}d\upsilon = \iiint_{\upsilon} \Delta V d\upsilon = \iint_{S} \frac{\partial V}{\partial n} dS. \tag{2.120}$$

Green's integral formulas are derived from Gauss' formula by specifying the components of **F** as:

$$F_X = U\frac{\partial V}{\partial X}, \tag{2.121}$$

where $X = x$, y or z and U and V are potentials.

Then it holds that:

$$\mathbf{F} \cdot \mathbf{n} = F_n = U\frac{\partial V}{\partial n}, \tag{2.122}$$

and from Eq. (2.119):

$$\mathrm{div}\mathbf{F} = U\Delta V + \frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z}, \tag{2.123}$$

so that *Green's formula I* is obtained from Eq. (2.118) as

$$\iiint_{\upsilon} \left(U\Delta V + \frac{\partial U}{\partial x}\frac{\partial V}{\partial x} + \frac{\partial U}{\partial y}\frac{\partial V}{\partial y} + \frac{\partial U}{\partial z}\frac{\partial V}{\partial z} \right) d\upsilon = \iint_{S} U\frac{\partial V}{\partial n} dS. \tag{2.124}$$

If U and V are interchanged in Eq. (2.124), one obtains another equation, which, subtracted from Eq. (2.124), yields *Green's formula II*

$$\iiint_{\upsilon} (U\Delta V - V\Delta U) d\upsilon = \iint_{S} \left(U\frac{\partial V}{\partial n} - V\frac{\partial U}{\partial n} \right) dS. \tag{2.125}$$

In the above equations, it is assumed that U and V and their first- and second-order derivatives are finite and continuous in the region υ.

Finally, specifying $U = 1/l$ in Green II yields *Green's formula III*

$$\iiint_{\upsilon} \frac{1}{l}\Delta V d\upsilon - pV = \iint_{S} \left[\frac{1}{l}\frac{\partial V}{\partial n} - V\frac{\partial}{\partial n}\left(\frac{1}{l}\right) \right] dS. \tag{2.126}$$

where:

$$p = \begin{cases} 4\pi, & if \quad P \quad inside \quad S \\ 2\pi, & if \quad P \quad on \quad S \\ 0, & if \quad P \quad outside \quad S. \end{cases} \tag{2.127}$$

Here P is the computation point.

Green III also holds if υ is the exterior space to the surface S and n is the interior normal to S, where:

$$p = \begin{cases} -4\pi, & if \quad P \quad outside \quad S \\ -2\pi, & if \quad P \quad on \quad S \\ 0, & if \quad P \quad inside \quad S. \end{cases} \tag{2.128}$$

Proof of Corollary 2.1 (Poisson's differential equation)

From Eq. (2.120) one obtains for point P located inside the closed masses inside volume ϑ:

$$\begin{aligned} \iiint_{\vartheta} \Delta V_P d\vartheta &= \iiint_{\vartheta_e} \Delta V_P d\vartheta = \varepsilon^2 \iint_{\sigma_e} \frac{\partial V}{\partial r} d\sigma \\ &= G\varepsilon^2 \iint_{\sigma_e} \frac{\partial}{\partial \varepsilon}\left(\frac{\rho}{\varepsilon}\right) d\sigma \to -4\pi G \rho_P \quad \text{as} \quad \varepsilon \to 0. \end{aligned}$$

Here ϑ_e and σ_e are the volume and surface of an infinitesimal sphere of radius ε centred at point P.

2.7.4.1 A Green's Formula on the Sphere

Meissl (1971, p. 12), with reference to Hotine (1969), presented the following integral relations derived from a Green's formula on the sphere

$$\iint_{\sigma} (\nabla^* f)^T \cdot \nabla^* g d\sigma = -\iint_{\sigma} f \Delta^* g d\sigma = -\iint_{\sigma} g \Delta^* f d\sigma, \tag{2.129a}$$

where ∇^* is *the gradient operator on the unit sphere*

$$\nabla^* = \left[\frac{\partial}{\partial \varphi} \quad \frac{\partial}{\cos\varphi \partial \lambda}\right]^T, \tag{2.129b}$$

which is related to the gradient operator on the sphere of radius R, ∇, by:

$$\nabla = \frac{1}{R}\nabla^* = \left[\frac{\partial}{\partial x} \quad \frac{\partial}{\partial y}\right]^T, \tag{2.129c}$$

where x and y are local horizontal coordinates on the sphere.

If one specifies $f = Y_{nm}$ and $g = Y_{pq}$, it follows from Eqs. (2.129a) and (2.117) that:

$$\iint_{\sigma} (\nabla^* Y_{nm})^T \cdot \nabla^* Y_{pq} d\sigma = -\iint_{\sigma} Y_{nm} \Delta^* Y_{pq} d\sigma = \begin{cases} 4\pi n(n+1), & if \quad (n,m) = (p,q) \\ 0 & otherwise. \end{cases} \tag{2.130}$$

i.e. also the gradients of the surface spherical harmonics are orthogonal on the sphere.

2.7.5 Boundary Value Problems

Stokes' theorem states that, for a given potential $V = V_S$ on a surface S, there is only one harmonic potential V in its exterior (if it exists).

Proof The proof follows from the following form of Green I

$$\iiint_{\upsilon} \left(U\Delta U + \left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial z}\right)^2 \right) d\upsilon = \iint_{S} U \frac{\partial U}{\partial n} dS. \tag{2.131}$$

Let us now assume that there are two potentials V_1 and V_2 in the exterior of S that take on the same values on S. Then the difference potential $U = V_1 - V_2$ on S and its Laplaceian ΔU in υ vanish, so that the integral reduces to:

$$\iiint_{\upsilon} \left(\left(\frac{\partial U}{\partial x}\right)^2 + \left(\frac{\partial U}{\partial y}\right)^2 + \left(\frac{\partial U}{\partial z}\right)^2 \right) d\upsilon = 0, \tag{2.132}$$

implying that U is a constant in υ. As U vanishes on S, it must vanish also outside S, and it follows that $V_1 = V_2$ in υ.

Stokes' theorem confirms that the forward (direct) problem in potential theory has a unique solution. On the other hand, the inverse problem (to determine the mass distribution that generates the external gravity field) is not unique.

Here follow some specific forward problems:

Dirichlet's (exterior) problem [or the first boundary value problem (bvp) of potential theory] is to determine the potential V outside the closed surface S, given the boundary values V_S. If the surface is a sphere, the solution is Poisson's integral formula for the sphere.

Neumann's (exterior) problem [or the second bvp of potential theory] is to determine V on and in the exterior of S from the given function $\partial V/\partial n$ on S. Here n is the exterior normal to the surface. If the surface is a sphere, the solution on the sphere is Hotine's formula and in the exterior it is Hotine's extended formula.

The third bvp is to determine the potential V in the exterior of S from boundary values $aV + b\partial V/\partial n$ on S, where a and b are constants. If the boundary values are

gravity anomalies, the 3rd bvp applies for determining the disturbing potential. If S is a sphere the solution is Stokes' formula, which is the basic formula for geoid determination from gravity anomalies. As we will see in Sect. 3.3, all three bvps are useful in physical geodesy.

2.8 Regularization

Geophysicists and physical geodesists are frequently confronted with linear-inverse problems, which can be solved in various ways. An *inverse problem* generally deals with the problem of converting observations $\tilde{g}$ to information w (of physical or other origin) that generates the observations. Frequently the problem is *ill-posed*, implying that the available (type of) observations are not sufficient to determine a unique solution for w. This can be illustrated by Poisson's integral formula in the exterior space of the sphere, Eq. (3.31a). Assuming that there are no masses outside the sphere of radius R, the forward problem to determine the disturbing potential or, in this case, the gravity anomaly Δg_P at any point P outside the sphere from gravity anomalies w on the sphere is given by the spectral solution

$$\Delta g_P = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+2} w_n(\theta, \lambda); \quad r_P > R \tag{2.133a}$$

where:

$$w_n(\theta, \lambda) = \sum_{m=-n}^{n} w_{nm} Y_{nm}(\theta, \lambda). \tag{2.133b}$$

Consider next that the gravity anomaly is known on an outer sphere of radius r_P, and the task is to solve the inverse problem of finding the anomaly on the lower sphere of radius R. Then we may develop Δg into spherical harmonics on the exterior sphere, yielding the coefficients Δg_{nm}, and, by comparing the spectral components with those in Eqs. (2.133a, b), one obtains the spectral equation and solution:

$$\Delta g_{nm} = \left(\frac{R}{r_P}\right)^{n+2} w_{nm} \Rightarrow w_{nm} = \left(\frac{r_P}{R}\right)^{n+2} \Delta g_{nm}, \tag{2.134a}$$

and formally the full solution becomes:

$$w(\theta, \lambda) = \sum_{n=0}^{\infty} \left(\frac{r_P}{R}\right)^{n+2} \Delta g_n(\theta, \lambda). \tag{2.134b}$$

In this case, as there are no masses between the spheres, the solution exists for erroneous observations, but in practice it will be severely ill-conditioned (more so in higher-degree harmonics), due to inevitable erroneous observations.

A *discrete ill-posed problem* occurs from the discretization of an ill-posed problem. A typical linear-inverse problem is that of estimating the density distribution or density structure inside the Earth from gravity or related data observed on or above the Earth's surface. Such a problem can frequently be expressed as a linear Fredholm integral equation of the 1st kind (e.g. Chambers 1976):

$$M\{K(P,Q)\tilde{w}\} = \tilde{g}(P), \tag{2.135a}$$

where $M\{\}$ is the integral over the surface of the Earth, or it is the mean value operator over the unit sphere (σ):

$$M\{\} = \frac{1}{4\pi}\iint_{\sigma} \{\} d\sigma, \tag{2.135b}$$

and $K(P,Q)$ is the kernel function that relates the observations $\tilde{g}$ at the observation point P and $\tilde{w}$ at the integration point Q. It goes without further discussion that solving an integral equation (inverse problem) is a much more difficult problem than that of just computing an integral formula expression (forward problem). This is particularly the case if the integral equation is ill-posed.

A more general and difficult problem is that of solving a non-linear integral equation. Such problems are treated in Ch. 8.

Below we will limit the discussion to that of solving Eq. (2.135a) in the case that the kernel function is *separable* in the form of a series of Legendre's polynomials $P_n(\cos\psi)$, i.e.

$$K(P,Q) = \sum_{n=0}^{\infty} (2n+1) k_n P_n(\cos\psi), \tag{2.136}$$

where ψ is the geocentric angle between the points P and Q. Inserting Eq. (2.136) into Eq. (2.135a) and interchanging summation and integration one obtains:

$$\sum_{n=0}^{\infty} k_n \tilde{w}_n(P) = \tilde{g}(P) = \sum_{n=0}^{\infty} \tilde{g}_n(P), \tag{2.137a}$$

where:

$$\left\{ \begin{matrix} \tilde{w}_n(P) \\ \tilde{g}_n(P) \end{matrix} \right\} = \frac{2n+1}{4\pi} M\left\{ \left[\begin{matrix} \tilde{w}(Q) \\ \tilde{g}(Q) \end{matrix} \right] P_n(\cos\psi) \right\} \tag{2.137b}$$

are the *Laplace harmonics* of $\tilde{w}$ and $\tilde{g}$. Although these harmonics are functions of position, below we will usually not specify this unless necessary for understanding.

From Eq. (2.137a), we may identify a relation between the unknown $\tilde{w}_n$ and the known $\tilde{g}_n$ as:

$$k_n \tilde{w}_n = \tilde{g}_n \quad \text{or} \quad \tilde{w}_n = \frac{\tilde{g}_n}{k_n} \quad \text{if and only if} \quad k_n \neq 0, \tag{2.138}$$

and these relations hold also for the error free harmonics w_n and g_n.

In this study, we will always assume that $k_n \neq 0$ for all degrees. Then, at least tentatively, one may come up with a solution for the unknown as:

$$\tilde{w}(P) = \sum_{n=0}^{\infty} \tilde{w}_n(P) = \sum_{n=0}^{\infty} \frac{\tilde{g}_n(P)}{k_n}. \tag{2.139}$$

However, this series does not necessarily converge, but in order to do so, $\tilde{g}_n$ must be smoother than k_n. More precisely, a square integrable solution for $\tilde{w}$ exists if only if the *Picard condition* is satisfied, i.e.

$$\sum_{n=0}^{\infty} \left(\frac{\tilde{g}_n}{k_n} \right)^2 < \infty. \tag{2.140}$$

This condition can be satisfied either by truncating the unknown spectrum of w to a finite degree, say, $n_{\max}$, by smoothing the coefficients k_n or both. In the first case, despite the truncation, the solution will be affected also by high-degree signals and the noise of the observations (spectral leakage; Trampert and Snieder 1996). In the second case, by discretizing Eq. (2.135a), one implicitly smooths the solution space to a finite set, corresponding to the selected block size on the sphere. Approximately, by choosing the block size $v^\circ \times v^\circ$, the resolution of the solution will be limited to harmonic degree $n_{\max} = 180/v^\circ$. The smaller the block size, the more ill-conditioned the system of equations will be. As an example, Martinec (1998, Sect. 8.6) performed a discrete downward continuation of surface gravity anomalies from elevations as high as 2.425 m to sea level in the Canadian Rocky Mountains by discretizing Poisson's integral equation Eq. (3.31a–c). The iterative solution worked well for observation grid sizes larger than or equal to 5′, while it failed for block sizes of 30″ × 60″ due to poor numerical conditioning.

Below we will study the solutions of Eq. (2.135a) by Tikhonov regularization, Wiener filter and spectral smoothing and combination. Other types of discrete regularization methods can be found in Hansen (1998).

2.8.1 Tikhonov Regularization

One method for regularization of an ill-posed problem originates with Phillips (1962) and AN Tikhonov in 1963 (see Tikhonov and Arsenin 1977). By this method, Eq. (2.135a) is first discretized into a matrix observation equation, where we assume that the system is over-determined, i.e. the number of observations is larger than the number of unknowns. The result is:

$$\mathbf{K}\tilde{\mathbf{w}} = \tilde{\mathbf{g}} - \boldsymbol{\varepsilon} = \mathbf{g}, \tag{2.141}$$

where $\mathbf{K}$, $\tilde{\mathbf{w}}$, $\tilde{\mathbf{g}}$ and $\boldsymbol{\varepsilon}$ are the design matrix, vectors of unknowns, observations and residuals, respectively. Assuming that the residuals are random with expectation zero, and that there are no correlations among the individual residuals, the related Tikhonov problem is to minimize the target function

$$E\left\{\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}\right\} + \mathbf{w}^T\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}\mathbf{w} \tag{2.142}$$

for some choice of the Tikhonov matrix $\boldsymbol{\Gamma}$. According to Ditmar et al. (2003), the problem of regularization includes two aspects: (a) the optimal choice of the regularization technique (i.e. of the regularizing functional or the regularization matrix) and (b) the optimal choice of the regularization parameter. The regularization matrix $\boldsymbol{\Gamma} = \alpha\mathbf{Q}$ can be divided into three categories: zero-order Tikhonov regularization with $\mathbf{Q} = \mathbf{I}$ and first- and second-order regularizations, where $\mathbf{Q}$ is either first- or second-order derivative operators (see Eqs. 2.143c, 2.143d). For $\boldsymbol{\Gamma} = \alpha\mathbf{I}$, where α is a small positive constant and $\mathbf{I}$ is the unit matrix, the solution to the minimization is given by the modified normal matrix equation

$$\left(\mathbf{K}^\mathbf{T}\mathbf{K} + \alpha^2\mathbf{I}\right)\hat{\mathbf{w}} = \mathbf{K}^\mathbf{T}\tilde{\mathbf{g}} \tag{2.143a}$$

with the solution

$$\hat{\mathbf{w}} = \left(\mathbf{K}^\mathbf{T}\mathbf{K} + \alpha^2\mathbf{I}\right)^{-1}\mathbf{K}^\mathbf{T}\tilde{\mathbf{g}}, \tag{2.143b}$$

where the matrix $\alpha^2\mathbf{I}$ stabilizes the original least squares solution obtained for $\alpha = 0$. As the stabilization has the less desired effect of making the solution biased, the size of α should be a compromise between the bias and the expected observation error propagation, and it must be sufficiently large to match the computer capacity to solve Eq. (2.143a). Higher-order Tikhonov regularization operators are given by (Hansen 1998, Chap. 8):

$$\mathbf{Q}_{(n-1)\times n} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \tag{2.143c}$$

and

$$\mathbf{Q}_{(n-2)\times n} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}, \tag{2.143d}$$

which represent the first- and second-derivative operators, respectively.

Applying singular value decomposition, matrix $\mathbf{K}$ can be decomposed into:

$$\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{2.144}$$

where $\mathbf{U}$ and $\mathbf{V}$ are matrices containing all the eigen-vectors $\mathbf{U}_i$ and $\mathbf{V}_i$ of $\mathbf{K}$, and $\mathbf{D}$ is a diagonal matrix constructed by the singular values (i.e., squares of the eigen-values λ_i of $\mathbf{K}$) As the eigen-vectors are orthonormal, it follows that Eq. (2.143a) has the solution

$$\hat{\mathbf{w}} = \mathbf{V}\left(\mathbf{D}^2 + \alpha^2\mathbf{I}\right)^{-1}\mathbf{D}\mathbf{U}^T\tilde{\mathbf{g}} = \sum_{i=1}^{q} \frac{\lambda_i^2 \mathbf{U}_i^T\tilde{\mathbf{g}}}{\lambda_i^2 + \alpha^2}\mathbf{V}_i = \sum_{i=1}^{q} f_i \frac{\mathbf{U}_i^T\tilde{\mathbf{g}}\mathbf{V}_i}{\lambda_i}, \tag{2.145}$$

where the filter factor $f_i = \lambda_i^2/\left(\lambda_i^2 + \alpha^2\right)$ smooths the solution for $\mathbf{w}$. By taking the statistical expectation of Eq. (2.145) and inserting the expected value for $\tilde{\mathbf{g}}$ from Eqs. (2.134a, b), it follows that each component $\hat{w}_i$ of the computed vector is biased by $-\alpha^2 w_i/\left(\lambda_i^2 + \alpha^2\right)$.

Note. Here we discuss only the simple Tikhonov regularization by Eq. (2.145). Other important methods can be found, e.g. in Hansen (1998, Sect. 5.1). In statistical literature Tikhonov's method is known as *ridge regression*, e.g. Marquardt (1970). Xu and Rummel (1994) presented such a technique, by introducing more than one regularization parameter, based on the criterion of minimizing the trace of the mean square error of the solution, to determine gravity potential harmonic coefficients from satellite gravimetric data.

Let us finally mention that one simple way of smoothing the Tikhonov type of solution is to limit the number of unknowns in Eq. (2.141), which corresponds to limiting the number of discrete surface elements in the integral of Eq. (2.135a), see for example the numerical study performed by Martinec (1998, Sect. 8.6).

2.8.1.1 Numerical Methods for Determining the Regularization Parameter α

There are some numerical methods for computing the regularization parameter α given in Eq. (2.145). Most important are the L-curve and generalized cross-validation methods (see Hansen 1998, Sect. 2). The L-curve displays the trade-off between minimizing the residual norm ($\|\boldsymbol{\varepsilon}\|_2 = \|\mathbf{KW} - \tilde{\mathbf{g}}\|_2$) and the solution norm ($\|\tilde{\mathbf{w}}\|_2$) in the regularization problem of Eq. (2.143a). For discrete ill-posed problems, it turns out that the L-curve, when plotted in logarithmic scale, almost always has a characteristic L-shaped appearance. Using this technique, we search for the point with maximum curvature as illustrated in Fig. 2.6. L-curve is the most suitable graphical tool to compute the regularization parameter, which is a plot for all regularization parameters of the norm of the regularized solution versus the corresponding residual norm. This method plays a major role in connection with regularization methods for discrete ill-posed problems.

Generalized cross-validation (GCV) is based on the assumption that, if the i-th arbitrary element of $\mathbf{w}$ is left out, then the corresponding regularized solution should predict this observation well, and the choice of regularization parameter should be independent of an orthogonal transformation of $\mathbf{w}$. Then the regularization parameter, given in Eq. (2.145), is chosen such that the following function is a minimum:

$$GCV = \frac{\|\mathbf{K}\hat{\mathbf{w}} - \tilde{\mathbf{g}}\|_2^2}{\left(trace(\mathbf{I}_m - \mathbf{K}\mathbf{K}^{-1})\right)} \tag{2.146}$$

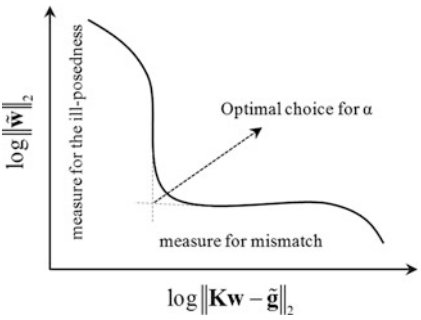

Fig. 2.6 The L-curve (graphical tool for analysis of discrete ill-posed problems)

2.8.2 Wiener Filtering

In the Wiener filter (Wiener 1949), one assumes that the signals g with w and the observation noise ε are all stationary stochastic processes with expectations zero with known spectral characteristics, i.e. their covariance and cross-covariance functions are known. The solution is practically the same as that for *least squares collocation* (Moritz 1980) and *kriging* (Matheron 1963) in geodesy and geostatistics, respectively.

Introducing a general estimator for w from a surface integral on sphere by the formula

$$\tilde{w}(P) = M_Q\{h(P,Q)\tilde{g}(Q)\}, \tag{2.147a}$$

where h is an arbitrary linear kernel function, the expected mean square error (MSE) the estimator becomes:

$$\begin{aligned}\tilde{m}^2 = E\left\{(\tilde{w}(P) - w(P))^2\right\} &= \sigma_w^2(P) - 2M_Q\left\{h(P,Q)c_{gw}(Q,P)\right\} \\ &+ M_Q\left[h(P,Q)M_X\left\{h(P,X)c_{\tilde{g}\tilde{g}}(Q,X)\right\}\right],\end{aligned} \tag{2.147b}$$

where c_{gw} and $c_{\tilde{g}\tilde{g}}$ are the cross- and auto-covariance functions between the signals marked by the subscripts, and $\sigma_w^2(P)$ is the variance of w. The minimum of the MSE is obtained for h satisfying the Wiener-Hopf equation (see also Sect. 2.2.2):

$$c_{wg}(P,Q) = M_X\left\{\hat{h}(P,X)c_{\tilde{g}\tilde{g}}(Q,X)\right\}, \tag{2.148}$$

yielding the MSE

$$\tilde{m}^2 = \sigma_w^2(P) - M_Q\left[\hat{h}(P,Q)M_X\left\{\hat{h}(P,X)c_{\tilde{g}\tilde{g}}(Q,X)\right\}\right]. \tag{2.149}$$

Assuming that the covariance functions are homogeneous and isotropic, they can be written in the spectral forms

$$c_{\tilde{g}\tilde{g}}(X,Q) = c_{gg}(X,Q) + c_{\varepsilon\varepsilon}(X,Q) = \sum_{n=0}^{\infty}\left(c_n^2 + \sigma_n^2\right)P_n(\cos\psi), \tag{2.150a}$$

and

$$c_{wg}(P,Q) = \sum_{n=0}^{\infty} d_n P_n(\cos\psi). \tag{2.150b}$$

Here c_n^2 and σ_n^2 are the signal and error degree variances of $\tilde{g}$, while $d_n = d_n(P)$ are the signal degree variances of the cross-covariance function. Notice that the

latter is a function of position/elevation (only) if the estimated quantity lies outside the sphere of computation.

Inserting the above series of covariance functions, as well as the series representation for the kernel function h,

$$h(P,Q) = \sum_{n=0}^{\infty} h_n P_n(\cos\psi) \tag{2.151}$$

into Eq. (2.148), the least squares solution for h follows as:

$$\hat{h}(P,Q) = \sum_{n=0}^{\infty} (2n+1)\frac{d_n}{c_n^2+\sigma_n^2} P_n(\cos\psi). \tag{2.152}$$

Hence, by considering Eqs. (2.147a) and (2.149), the least squares solution for w and its mean square error become:

$$\hat{w}(P) = \sum_{n=0}^{\infty} \frac{d_n}{c_n^2+\sigma_n^2}\tilde{g}_n \tag{2.153a}$$

and

$$\hat{m}^2 = \sigma_w^2 - \sum_{n=0}^{\infty} \frac{d_n^2}{c_n^2+\sigma_n^2}. \tag{2.153b}$$

From Eq. (2.138), we also have the relation $w_n = g_n/k_n$, which yields:

$$\hat{w}(P) = \sum_{n=0}^{\infty} f_n k_n^{-1}\tilde{g}_n, \quad \text{where } f_n = \frac{c_n^2}{c_n^2+\sigma_n^2} \tag{2.154a}$$

and

$$\hat{m}^2 = \sum_{n=0}^{\infty} k_n^{-2}c_n^2 - \sum_{n=0}^{\infty} k_n^{-2}\frac{c_n^4}{c_n^2+\sigma_n^2} = \sum_{n=0}^{\infty} k_n^{-2}\frac{c_n^2\sigma_n^2}{c_n^2+\sigma_n^2}. \tag{2.154b}$$

As an alternative, Eq. (2.154a) can be written in the space domain as:

$$\hat{w}(P) = \frac{1}{4\pi}\iint_{\sigma} H(P,Q)\tilde{g}(Q)d\sigma_Q, \tag{2.155a}$$

where the kernel function is:

$$H(P,Q) = \sum_{n=0}^{\infty} (2n+1) k_n^{-1} f_n P_n(\cos\psi). \tag{2.155b}$$

2.8.3 Spectral Smoothing

Let us return to Eq. (2.154a)

$$\tilde{w}(P) = \sum_{n=0}^{\infty} f_n k_n^{-1} \tilde{g}_n, \tag{2.156}$$

where f_n are now arbitrary filter parameters to be estimated such that the mean square error of $\tilde{w}$ is minimized. As the error of $\tilde{w}$ is given by its random error and bias, i.e.

$$\varepsilon_{\tilde{w}} = \sum_{n=0}^{\infty} \left[f_n k_n^{-1} \varepsilon_n + (f_n - 1) k_n^{-1} g_n \right], \tag{2.157}$$

it follows the expected MSE becomes:

$$\tilde{m}^2 = E\left[M\left\{ \varepsilon_{\tilde{w}}^2 \right\} \right] = \sum_{n=0}^{\infty} k_n^{-2} \left[f_n^2 \sigma_n^2 + (f_n - 1)^2 c_n^2 \right], \tag{2.158}$$

where σ_n^2 and c_n^2 are the error and signal-degree variances of $\tilde{g}$. The minimum MSE is achieved by differentiating the MSE w.r.t. each of the smoothing factors and equating to zero. The result is:

$$\hat{f}_n = \frac{c_n^2}{c_n^2 + \sigma_n^2}, \tag{2.159}$$

and the least squares estimator $\hat{w}$ and its MSE are the same as in Eqs. (2.154a, b). Some of the theory and applications were presented in Sjöberg (1980, 1986, 2011a, b).

The practical formulation in the space domain again becomes Eqs. (2.154a, b).

2.8.4 Spectral Combination

We now assume that, in addition to the information given in above, there is an Earth Gravitational Model (EGM) available to degree $n_{\max}$ that yields the unbiased estimate w_1 (unbiased through degree $n_{\max}$)

$$w_1 = \sum_{n=0}^{n_{\max}} w_n^{GM}, \tag{2.160}$$

with the random error with expectation zero

$$dw_1 = \sum_{n=0}^{n_{\max}} w_n^{GM} \tag{2.161}$$

and the variance (composed of the error degree variances σ_n^{GM})

$$\sigma_{w1}^2 = \sum_{n=0}^{n_{\max}} \sigma_n^{GM}. \tag{2.162}$$

A general estimator for w, unbiased through degree $n_{\max}$, can be formulated as:

$$\tilde{w} = \sum_{n=0}^{\infty} k_n^{-1} p_n \tilde{g}_n + \sum_{n=0}^{n_{\max}} (1 - p_n) w_n^{GM}, \tag{2.163}$$

with the MSE

$$\tilde{m}^2 = \sum_{n=0}^{\infty} \left\{ k_n^{-2} p_n^2 \sigma_n^2 + (1 - p_n)^2 dc_n^{GM} \right\}, \tag{2.164a}$$

where:

$$dc_n^{GM} = \begin{cases} \sigma_n^{GM} & if \quad n \le n_{\max} \\ k_n^{-2} c_n^2 & otherwise. \end{cases} \tag{2.164b}$$

Here p_n are arbitrary degree weights, which are optimized in a least squares sense by differentiating $\tilde{m}^2$ w.r.t. each of them and equating to zero. The result is:

$$\hat{p}_n = \begin{cases} \frac{\sigma_n^{GM}}{k_n^{-2}\sigma_n^2 + \sigma_n^{GM}} & if \quad n \le n_{\max} \\ \frac{c_n^2}{c_n^2 + \sigma_n^2} & otherwise, \end{cases} \tag{2.165}$$

and the MSE becomes:

$$\hat{m}^2 = \sum_{n=0}^{n_{\max}} \frac{k_n^2 \sigma_n^2 \sigma_n^{GM}}{\sigma_n^2 + k_n^2 \sigma_n^{GM}} + \sum_{n=n_{\max}+1}^{\infty} \frac{k_n^2 c_n^2}{\sigma_n^2 + k_n^2 c_n^2}. \tag{2.166}$$

Finally, the spectral combination can be formulated also as the sum of a surface integral and a spectral series:

$$\hat{w}(P) = \frac{1}{4\pi} \iint_{\sigma} K(P,Q)\tilde{g}(Q)d\sigma_Q + \sum_{n=0}^{n_{\max}} (1-\hat{p}_n)^2 w_n^{GM}, \tag{2.167a}$$

where the kernel function is:

$$K(P,Q) = \sum_{n=0}^{\infty} (2n+1))k_n^{-1}\hat{p}_n P_n(\cos\psi). \tag{2.167b}$$

For applications of spectral smoothing and combination, see Sects. 4.4.4 and 7.5. See also Sjöberg (1981, 1986, 2011a, b).

2.8.5 Optimum Regularization

Based on the above experiences, one may ask whether Tikhonov regularization can be modified to share the properties of the Wiener filter and/or spectral smoothing, namely to be optimum in the sense of minimizing the MSE. There are numerous publications solving Tikhonov's regularization problem by minimizing the MSE of the solution. However, each such solution is optimal only w.r.t. the specified target function, Eq. (2.142), i.e. for a specified Tikhonov matrix $\mathbf{\Gamma}$. Hence, the major problem is thus to find the correct Tikhonov matrix for the optimum solution. For this purpose, we rewrite Eq. (2.135a) as:

$$M\{[K(\psi) + Q(\psi)]\tilde{w}\} = \tilde{g}(P), \tag{2.168a}$$

where:

$$Q(\psi) = \sum_{n=0}^{\infty} (2n+1)q_n P_n(\cos\psi). \tag{2.168b}$$

Here q_n are arbitrary parameters to be determined such that the target function, the expected global MSE m_w^2 of $\tilde{w}$, is minimized. As the spectral form of Eq. (2.168a) can be written:

$$\tilde{w}_n = \frac{\tilde{g}_n}{k_n + q_n}, \tag{2.169}$$

it follows that its error and global MSE become:

$$\varepsilon_{w_n} = \frac{\tilde{g}_n}{k_n + q_n} - \frac{g_n}{k_n} \quad \text{and} \quad m_{w_n}^2 = \frac{k_n^2\sigma_n^2 + q_n^2 c_n^2}{k_n^4 + k_n^2 q_n^2}, \tag{2.170}$$

and, by differentiating the MSE w.r.t. q_n, one obtains the least squares choice of the parameters:

$$\hat{q}_n = k_n \sigma_n^2 / c_n^2. \tag{2.171}$$

Inserting this choice for q_n in Eq. (2.169) and summing up, one obtains the solution

$$\hat{w} = \sum_{n=0}^{\infty} \frac{c_n^2}{c_n^2 + \sigma_n^2} \frac{\tilde{g}_n}{k_n}, \tag{2.172}$$

which is the same as the solution by spectral filtering, Eqs. (2.156), with filter factors given by Eq. (2.159).

However, the kernel function Q with parameters $\hat{q}_n$ is a divergent series, as $\sigma_n^2 / c_n^2 > 1$ for large n, which implies that the optimum regularization cannot be realized in the limiting integral equation, Eq. (2.168a). However, in the numerical approximation of the integral equation, we may approximate it by the matrix equation

$$\left(\mathbf{K} + \overline{\mathbf{Q}}\right)\mathbf{w} = \tilde{\mathbf{g}} \quad \text{(consistent)} \tag{2.173}$$

with the solution

$$\mathbf{w} = \left(\mathbf{K} + \overline{\mathbf{Q}}\right)^{-1} \tilde{\mathbf{g}}, \tag{2.174}$$

where the elements of $\overline{\mathbf{Q}}$ are determined from a smoothed kernel function, Eq. (2.168b), e.g. obtained by truncating the series to a maximum degree. (Such a truncation is a reasonable approximation, as the numerical integration to a finite number of integration blocks will automatically limit the frequency contained in the solution.) In this way, the solution for $\mathbf{w}$ will be a smoothed spectral filter/Wiener filter. The higher the degree of truncation in the kernel function for $\overline{\mathbf{Q}}$, and the more precise the numerical integration is, the closer to the Wiener filter will be the solution.

In the special case with all q_n set to a constant α^2, the function value for $\mathbf{Q}(0)$ is still infinite, but the Tikhonov solution is consistent with the non-optimized Wiener filter

$$\hat{w} = \sum_{n=0}^{\infty} \frac{c_n^2}{c_n^2 + \alpha^2} \frac{\tilde{g}_n}{k_n}. \tag{2.175}$$

2.8.6 Spherical Harmonic Analysis

Let the unknown w be represented by the truncated series of fully normalized spherical harmonics Y_{nm}:

$$\tilde{w}(P) = \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} w_{nm} Y_{nm}(P), \tag{2.176a}$$

where the harmonics are orthonormal, i.e.

$$M\{Y_{nm} Y_{rs}\} = \begin{cases} 1, & \text{if} \quad (n,m) = (r,s) \\ 0 & \text{otherwise.} \end{cases} \tag{2.176b}$$

The task is to determine the spherical harmonic coefficients w_{nm} from Eq. (2.168a). Hence, by inserting Eq. (2.176a) and using Eq. (2.176b), we obtain:

$$\sum_{n=0}^{n_{\max}} (k_n + q_n) \sum_{m=-n}^{n} w_{nm} Y_{nm}(P) = \tilde{g}(P). \tag{2.177}$$

Then, by multiplying each member of this equation by $Y_{nm}(P)$ and averaging over the unit sphere, one finally arrives at the solution for the harmonic coefficients

$$w_{nm} = \frac{1}{k_n + q_n} M\{\tilde{g}(P) Y_{nm}(P)\}. \tag{2.178}$$

If q_n is chosen as $k_n \sigma_n^2 / c_n^2$, this solution for the harmonic coefficients will be optimal in the sense of minimizing the MSE. Finally, by applying these coefficients in Eq. (2.176a), the optimum, truncated estimate of w is obtained.

The solution by spherical harmonic analysis has the merits of being stable and not prone to spectral leakage as previous methods (see the different approach of Trampert and Snieder (1996) and Spetzler and Trampert (2003), which suffers from the problem of leakage). Its major drawback is the requirement of a global, homogeneous coverage of data on the sphere. Finally, we mention that if an independent set of harmonics of w is available, it can be combined with the above harmonics in an optimum sense.

This method was applied [with q_n set to 0 in Eq. (2.178)] in computing some of the OSU Earth Gravitational Models in the 1970s and 1980s; see e.g. Rapp (1981) and Rapp and Cruz (1986).

2.8.7 Comparison

Table 2.1 summarizes a comparison of the inverse solutions by Tikhonov's method, Wiener filtering and spectral smoothing and combination.

The most important results are the following:

- Tikhonov regularization is the solution by direct (approximate) solving the original integral equation. The smoothing is performed in two ways: (1) the original integral equation is approximated by a finite sum of unknowns and surface elements, and (2) by adding the smoothing term $\alpha^2\mathbf{I}$ to the normal matrix. All other methods use a direct integral formula for the solution.
- Although the spectral solution of Tikhonov regularization (Eq. 2.145) resembles the spectral forms of the Wiener filter and spectral smoothing, they are not the same, as the Tikhonov solution includes the singular values and eigen-vectors of the normal matrix (including the smoothing term), which vary w.r.t. chosen block size and number of unknown parameters, while the latter are based on the kernel, signal and error spectra. This implies that the individual observations used in Tikhonov's method can be weighted, but the additional feature of spectral weighting in the other methods, is not possible.
- The Wiener filter and the spectral smoother are identical solutions. However, the assumptions and target functions differ. The former minimizes the (local) MSE based on the known signal and covariance functions (correlation functions), while the latter minimizes the global MSE base on known signal and error degree variances.

Table 2.1 Comparison of regularization methods (Sjöberg 2012)

Method	Equation	Extra info.	Target function	Assumptions
Tikhonov regularization	Integral equation	α	$\boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon} + \alpha^2\mathbf{I}$	$E\{\varepsilon\} = 0$
Wiener filter	Integral formula	c_n^2, σ_n^2, d_n	MSE	Stochastic processes[a]
Spectral smoothing	Integral formula	c_n^2, σ_n^2, d_n	Global MSE	Stochastic processes[b]
Spectral combination	Integral formula	c_n^2, σ_n^2, d_n	Global MSE	Stochastic processes[b]
Optimum regularization	Integral equation	c_n^2, σ_n^2, d_n	Global MSE	Stochastic processes[c]
Harmonic analysis	Integral formula	c_n^2, σ_n^2	MSE	Global data on the sphere

[a]The stochastic processes are stationary processes with expected value zero
[b]The expected values of the observation errors are zero
[c]The optimum Q is approximated by the smoothed $\overline{\mathbf{Q}}$

- The Wiener filter assumes stationary stochastic processes with statistical expectations zero, while spectral smoothing and combination only assumes that the observation errors are zero in expectations.
- By spectral combination the integral solution can be combined in an optimum way with a priori information of the unknown function in the spectral domain.
- The solution by optimum regularization, described in Sect. 2.8.5, is the space domain representation of spectral smoothing. The solution is a smoothed spectral/Wiener filter. In the continuous case, the problem cannot even be formulated, as the corresponding integral equation does not exist.
- Harmonic analysis needs data from all over the sphere. Among its merits are resistance to spectral leakage and ill-conditioning.

2.8.8 *Concluding Remarks*

As discussed above, the numerical solution to an integral equation is frequently ill-conditioned, and for ill-posed problems this is always the case. Then, a unique and stable solution can be obtained, at the prize of a bias, by introducing some kind of smoothing. The numerical Tikhonov types of solutions involve solving a matrix equation, where the biases are based on some criterions. In contrast, the solutions by Wiener filter and spectral smoothing and combination are more computationally efficient, as they employ forward integration, or, numerically, matrix multiplications. This implies that Tikhonov types of solutions are particularly sensitive to the choice of block-size in the numerical integration of the coefficients of the design matrix, and the bias term α^2 must increase when the block-size decreases. This problem is not the case for the direct integration methods of Wiener and spectral filtering and spectral combination, as well as harmonic analysis.

Finally we emphasize that spectral combination is more flexible than the other methods, as it enables an optimal (with respect to minimum MSE) merging of different data. Also, harmonic analysis (possibly including spectral combination) is a viable alternative, provided that the data can be made available globally on the sphere.

Appendix: Answers to Exercises

Exercise 2.1 From Eq. (2.37a) one obtains the Taylor series

$$\begin{aligned} l_{P0}^{-1} &= r_P^{-1}\left[1-\frac{1}{2}\left(s^2-2st\right)+\frac{\left(-\frac{1}{2}\right)\left(-1-\frac{1}{2}\right)}{1\times 2}\left(s^2-2st\right)^2+\cdots\right] \\ &= r_P^{-1}\left[1+st+s^2\frac{3t^2-1}{2}+\cdots\right], \end{aligned}$$

and by comparing with Eq. (2.38a) the solution follows.

Exercise 2.2 The left member of Eq. (2.42) yields for $n = 0$, 1 and 2:

$$\int_{-1}^{1} 1dt = 2, \int_{-1}^{1} t^2 dt = 2/3 \quad \text{and} \quad \int_{-1}^{1} \left(\frac{3t^2-1}{2}\right)^2 dt$$
$$= \int_{-1}^{1} \left(\frac{9t^4-6t^2+1}{4}\right) dt = \frac{1}{4}\left[\frac{9t^5}{5} - \frac{6t^3}{3} + t\right]_{-1}^{1} = \frac{2}{5},$$

The second equation in Eq. (2.42) is shown in the same way.

References

Bjerhammar A (1973) Theory of errors and generalized matrix inverses. Elsevier, Amsterdam

Bjerhammar A (1975) Discrete approaches to the solution of the boundary value problems of physical geodesy. Boll Geod Szi Aff 34(2):185–240

Chambers LG (1976) Integral equations—a short course. Int Text Book Co Ltd, London

Danby JMA (1988) Fundamentals of celestial mechanics. Willmann-Bell, Richmond

Dermanis A (1984) Kriging and collocation—a comparison. Manusc Geod 9(3):159–167

Fukushima T (2006) Transformation from Cartesian to geodetic coordinates accelerated by Halley's method. J Geodesy 79:689–693

Ditmar P, Kusche J, Klees R (2003) Computation of spherical harmonics coefficients from gravity gradiometry data to be acquired by the GOCE satellite: regularization issues, J Geod 77:465–477

Gleason DM (1988) Comparing ellipsoidal corrections to the transformation between the geopotential's spherical and ellipsoidal spectrums. Manuscr Geod 13:114–129

Hansen PC (1998) Rank-deficient and discrete ill-posed problems. Siam, Philadelphia

Heiskanen W, Moritz H (1967) Physical geodesy. W.H. Freeman and company, San Francisco

Hofmann-Wellenhof B, Lichtenegger H, Wasle E (2008) GNSS global navigation satellite systems; GPS, Glonass, Galileo & more. Springer, New York, 501pp

Hotine M (1969) Mathematical Geodesy, ESSA Monograph 2. U.S. Department of Commerce, Washington, D.C.

Krarup T (1969) A contribution to the Mathematical Foundation of Physical Geodesy, Meddelelse No. 44, Geodaetisk Institut, Copenhagen

Lauritzen SL (1973) The probabilistic background of some statistical methods in physical geodesy. Danish Geodetic Inst., Meddelelse No. 48, Copenhagen

Marquardt DW (1970) Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12:591–612 (American Society for Quality and the American Statistical Association)

Martinec Z (1998) Boundary–value problems for gravimetric determination of a precise geoid. Lecture notes in earth sciences. Springer, Berlin, Heidelberg

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Meissl P (1971) Preparation for the numerical evaluation of second order Molodensky-type formulas. Dept Geodetic Science Rep No. 163, Ohio State Univ, Columbus, Ohio

Moritz H (1980) Advanced physical geodesy. H Wichmann, Karlsruhe

Pavlis NA, Simon AH, Kenyon SC, Factor JK (2012) The development and evaluation of the earth gravitational model 2008 (EGM2008). JGR 117:B04406

Phillips DL (1962) A technique for the numerical solution of certain integral equations of the first kind. J Assoc Comput Mach 9:84–97
Rapp RH (1981) On the Earth's gravity field to degree and order 180 using Seastat altimeter data, terrestrial gravity data and OAU data. Report No. 367, Dept Geod Sci and Surv, The OSU, Columbus, OH
Rapp RH, Cruz JY (1986) The representation of the Earth's gravitational potential in a spherical
Sjöberg L (1975) On the discrete boundary value problem of physical geodesy with harmonic reduction to an internal sphere. Ph.D. thesis, Dept. of Geod., Royal Inst. of Tech., Stockholm
Sjöberg L (1978) The accuracy of geoid undulations by degree implied by mean gravity anomalies on a sphere. OSU, Columbus and The Royal Institute of Technology, Stockholm, May 1978; J Geophys Res 84(B11):6226–6230
Sjöberg LE (1979) Integral formulas for heterogeneous data in physical geodesy. Royal Institute of Technology, Stockholm, Nov 1978. Bull Geodesique 53:297–315
Sjöberg LE (1980) Least squares combination of satellite harmonics and integral. Formulas in physical geodesy. Gerlands Beitr Geophys 895:371–377
Sjöberg LE (1981) Least squares combination of satellite and terrestrial data in physical geodesy. Ann Geophys 371:25–30
Sjöberg LE (1984a) Least squares modification of Stokes' and Vening Meinesz' formulas by accounting for truncation and potential coefficient errors. Manusc Geod 9:209–229
Sjöberg LE (1984b) Least squares modification of Stokes' and Vening Meinesz' formulas by accounting for errors of truncation, potential coefficients and gravity data. Dept Geod Report No. 27, University of Uppsala, 16
Sjöberg LE (1986) Comparison of some methods of modifying Stokes' formula. Boll Geod Scie Aff 453:229–248
Sjöberg LE (1999) An efficient iterative solution to transform rectangular to geocentric coordinates. Z Vermessungswesen 126(9):295–297
Sjöberg LE (2005) A local least-squares modification of Stokes' formula. Stud Geophys Geod 49:23–30
Sjöberg LE (2008) A strict transformation from Cartesian to geodetic coordinates. Surv Rev 40(308):156–163
Sjöberg LE (2011a) Geoid determination by spectral combination of an Earth gravitational model with airborne and terrestrial gravimetry. Stud Geophys Geodaet 55(4):57
Sjöberg LE (2011b) Local least squares spectral filtering and combination by harmonic functions on the sphere. J Geod Sci 1:355–360
Sjöberg LE (2012) Solutions to linear inverse problems on the sphere by Tikhonov regularization, Wiener filtering and spectral combination—a comparison. J Geod Sci 2:31–37
Sjöberg LE (2013) Theory of satellite geodesy. Division of Geodesy, Royal Institute of Technology, 143pp (Lecture Notes)
Spetzler J, Trampert J (2003) Implementing spectral leakage corrections in global surface wave tomography. Geophys J Int 155:532–538
Tarantola A (1987) Inverse problem theory. Methods for data fitting and model parameter estimation. Elsevier, Amsterdam
Tikhonov AN, Arsenin VY (1977) Solutions of ill-posed problems. Winston, Washington, D.C.
Trampert J, Snieder R (1996) Model estimations biased by truncated expansions: possible artefacts in seismic tomography. Science 271:1257–1260
Wenzel H-G (1981) Zur Geoidbestimmung durch Kombination von Schwereanomalien und einem Kugelfunktionsmodell mit Hilfe von Integralformeln. ZfV 1063:102–111
Wiener N (1949) Extrapolation, interpolation and smoothing of stationary time series. Wiley, New York. ISBN 0-262-73005-7
Xu P, Rummel R (1994) Generalized ridge regression with applications in determination of potential fields. Manuscr Geod 20:8–20

Chapter 3
Classical Physical Geodesy

Abstract This chapter presents the basics of classical physical geodesy, starting from the definitions of the gravity potential and gravity. The normal gravity potential is derived as the potential of a level ellipsoid plus the rotational potential of the ellipsoid. Normal gravity, defined as the gradient of the normal gravity potential, is presented on and above the level ellipsoid. The basic concepts of the geoid, reference ellipsoid, disturbing potential and geoid height are defined, as well as the classical definitions of gravity anomaly and disturbing potential. After derivation of the fundamental equation of physical geodesy, the gravity field components of the disturbing potential, gravity anomaly and its radial derivative are presented in spherical harmonics, followed by Kaula's power rule of the geopotential harmonics. The classical integral formulas of Poisson, Stokes, Hotine, Vening Meinesz and the vertical gradient of gravity anomaly are derived by spherical harmonics. Other spherical integral formulas are derived for determining the gravity anomaly and/or disturbing potential from deflections of the vertical (inverse Vening Meinesz formula) and gravity gradient components. The classical procedures in geoid determination, including direct and secondary indirect topographic effects and downward continuation of gravity and primary indirect topographic effect on the disturbing potential, are described. Finally, the chapter deals with common height systems, such as geopotential numbers, dynamic, orthometric and normal heights, as well as normal-orthometric heights. Some approximate formulas to correct normal-orthometric heights to orthometric or normal heights are also presented.

Keywords Disturbing potential · Gravity anomaly · Height systems · Integral formulas

3.1 Introduction

Most geodetic observations depend in one way or another on the Earth's gravity field, and both geodesists and geophysicists make use of geodetic data. Hence, a basic understanding of physical geodesy is a natural part of a geodesist's education and broadens the knowledge of the geophysicist.

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_3

Classical physical geodesy deals with handling and converting gravity data into gravity anomalies and disturbances and applying them to estimate the geoid, deflections of the vertical and gravity gradients. Also, geodetic height systems and their determination belong to physical geodesy. Classical literature on the subject are Heiskanen and Moritz (1967) and, to some extent, Heiskanen and Vening Meinesz (1958).

3.2 Basic Concepts in Physical Geodesy

3.2.1 The Gravity Field

The Earth's *gravity potential* (or *geopotential*) W is the sum of its gravitational potential V and rotational potential $\overline{\Omega}$:

$$W = V + \overline{\Omega}. \tag{3.1}$$

The gravitational potential is generated by the attraction of all the masses of the Earth and its atmosphere, and

$$\overline{\Omega} = \frac{\omega^2}{2}\left(x^2 + y^2\right) = \frac{\omega^2}{2} r^2 \sin^2\theta \tag{3.2}$$

depends on the Earth's angular velocity (ω) and the position of the observer in the (x, y)-plane of the global geodetic reference system. Here r is the radius from the geocentre and θ is the co-latitude (see Fig. 3.1.) The centrifugal force is the gradient of the rotational potential.

Exercise 3.0 Show that the geopotential is not harmonic even in the exterior of the Earth.

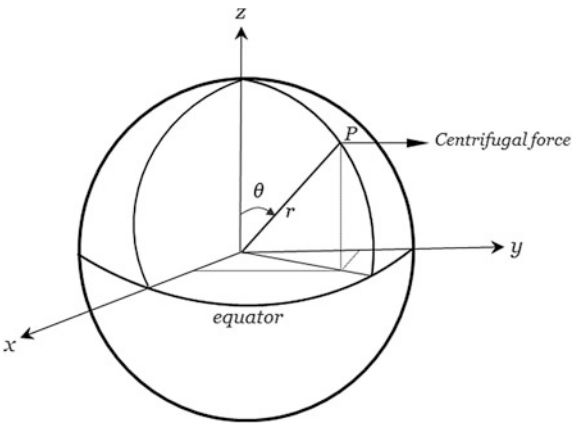

Fig. 3.1 The centrifugal force

Equation (3.1) implies that the gravity vector can be written:

$$\mathbf{g} = \begin{pmatrix} g_x & g_y & g_z \end{pmatrix}^T = -grad(W) = -grad(V) - grad(\Omega), \tag{3.3}$$

i.e. it is the sum of the gravitational and centrifugal force vectors but with opposite sign. (The sign is due to convention such that the gradient is directed towards the attracting masses.) The magnitude of gravity

$$g = \sqrt{g_x^2 + g_y^2 + g_z^2} \tag{3.4}$$

varies between 978 Gal at the poles and 983 Gal at the Equator. This difference is due to the flattening of the Earth towards the poles. This yields a gravity flattening of (983 − 978)/983 = 1/197, while the Earth's geometric flattening is 1/298 (see Sect. 1.5). The gravitational force unit Gal, named after the Italian astronomer G. Galileo, is defined as 1 Gal = 1 cm/s^2. In practical physical geodesy, but also in gravimetry and geophysics, the unit mGal is more practical.

Exercise 3.1 Assume that the magnitude of the gravitational vector is 983 Gal at the surface of a spherical Earth model with radius 6371 km. Where on the model are the maximum and minimum gravity and what is the difference?

Exercise 3.2 Consider a point at 60° latitude on the Earth model above. How much would the direction of the plumb-line change and in what direction if the Earth's rotation stops?

Solutions to Exercises are given in Appendix 2.

3.2.2 The Gravity Field of the Level Ellipsoid

A level ellipsoid is an ellipsoid with constant surface potential. The gravity field of the level ellipsoid is practically important, as it is used as *the normal gravity field*. By subtracting the normal potential and normal gravity from the Earth's potential and gravity (both at the same point in the space), one obtains the much smaller quantities called the disturbing potential and gravity disturbance, respectively.

- *The Normal Potential*

The normal gravity field is that generated by a level ellipsoid ($U_0 = \text{constant}$), and it is defined as the sum of the rotation potential $\overline{\Omega}$ (being the same as for the actual geopotential) and the gravitational potential (denoted X below). As the level ellipsoid is symmetric around the rotation axis, its gravitational potential can be expressed as a series in ellipsoidal harmonics of Legendre's polynomials in terms of only even degrees, yielding the normal potential:

$$U(u,\beta) = X + \overline{\Omega} = \sum_{n=0}^{\infty} \frac{Q_{2n}(iu/\mathrm{E})}{Q_{2n}(ib/\mathrm{E})} X_{2n} P_{2n}(\sin\beta) + \frac{\omega^2(u^2+E^2)}{2}\cos^2\beta, \quad (3.5)$$

and on the level ellipsoid with $u = b$ the formula reduces to:

$$U(b,\beta) = U_0 = \sum_{n=0}^{\infty} X_{2n} P_{2n}(\sin\beta) + \frac{a^2\omega^2}{2}\cos^2\beta. \quad (3.6)$$

However, as the normal potential is constant on the ellipsoid, implying that it is independent of β, and $\cos^2\beta = 2[1 - P_2(\sin\beta)]/3$, it follows that all coefficients X_{2n} of degrees other than zero and two vanish, $V_0 = U_0 - V_2$ and $V_2 = a^2\omega^2/3$. As a result, the gravitational potential of the normal potential (the sum in Eq. 3.5) becomes:

$$X(u,\beta) = \left(U_0 - \frac{a^2\omega^2}{3}\right)\frac{Q_0(iu/E)}{Q_0(ib/E)} + \frac{a^2\omega^2}{3}\frac{Q_2(iu/E)}{Q_2(ib/E)}P_2(\sin\beta), \quad (3.7a)$$

where

$$Q_0(iu/E) = -i\,\arctan(E/u) \quad (3.7b)$$

and

$$Q_2(iu/E) = \frac{i}{2}\left[\left(1+3\frac{u^2}{E^2}\right)\arctan\left(\frac{E}{u}\right) - 3\frac{u}{E}\right] = iq(u). \quad (3.7c)$$

For large distances outside the ellipsoid one has:

$$\arctan\left(\frac{E}{u}\right) \approx \frac{E}{u} \approx \frac{E}{r},$$

which yields:

$$X \approx \left(U_0 - \frac{\omega^2 a^2}{3}\right)\frac{\mathrm{E}}{\arctan(E/b)}\frac{1}{r} = \frac{GM}{r}.$$

The last equation implies that:

$$U_0 = \frac{GM}{E}\arctan(E/b) + \frac{\omega^2 a^2}{3},$$

so that one finally obtains:

$$U(u,\beta) = \frac{GM}{E}\arctan(\mathrm{E/u}) + \frac{\omega^2 a^2}{3}\frac{q(u)}{q(b)}P_2(\sin\beta) + \frac{\omega^2(u^2+E^2)}{2}\cos^2\beta. \quad (3.8)$$

This formula shows that the normal potential is given by the four parameters M, ω, a and E, or M, ω, a and b, that is the mass and angular velocity around the symmetry axis and semi axes of the reference ellipsoid.

The first two terms in Eq. (3.8) yield the gravitational potential of the level ellipsoid. It can be expressed in a series of spherical harmonics (see Heiskanen and Moritz 1967, Sects. 2–9):

$$X(r,\theta) = \frac{GM}{r}\left[1 - \sum_{n=1}^{\infty} X_{2n}\left(\frac{a}{r}\right)^{2n} P_{2n}(\sin\theta)\right], \quad (3.9a)$$

where

$$X_{2n} = (-1)^{n+1}\frac{3e^{2n}}{(2n+1)(2n+3)}\left(1 - n + 5n\frac{C-A}{ME^2}\right). \quad (3.9b)$$

Here C is the moment of inertia w.r.t. the axis of rotation, and A is the moment of inertia w.r.t. any axis in the equatorial plane. As X_{2n} is of the same order as the first eccentricity $(e = E/a)$ squared, it reduces quickly with the degree.

All the constants GM, a, e^2, ω and J_2 are specified in each geodetic reference system, such as GRS1980 (Sect. 1.5), which means that the normal potential can be conveniently determined everywhere on and outside the ellipsoid, either in a closed form by Eq. (3.8) or as a harmonic series for the normal gravitational potential by Eqs. (3.9a) and (3.9b).

- *Normal Gravity*

Normal gravity (γ) is given by:

$$\gamma = -grad\ U = -\frac{\partial U}{\partial h} = -\frac{1}{s_u}\frac{\partial U}{\partial u}, \quad (3.10a)$$

where

$$s_u = \sqrt{\frac{u^2 + E^2\sin^2\beta}{u^2+E^2}} \quad (3.10b)$$

is the line element for u, i.e. the change du yields the distance change $s_u du$, and h is the geodetic height, which is along the normal to the reference ellipsoid. Normal gravity on the ellipsoid (γ_0) can thus be obtained from Eqs. (3.10a) and (3.10b) by setting $u = b$. It can be written in a closed form as (Heiskanen and Moritz 1967, Eq. 2.69)

$$\gamma_0 = \frac{GM}{a^2\Psi}\left[1 + me'\frac{q'(b)}{q(b)}\left(\frac{\sin^2\beta}{2} - \frac{1}{6}\right) - m\cos^2\beta\right] \tag{3.11a}$$

where $\Psi = \sqrt{1 - e^2\cos^2\beta}$, $e' = E/b$, $m = \omega^2 a^2 b/(GM)$ and, with $q(u)$ defined as in Eq. (3.7c),

$$q'(u) = -\frac{u^2 + E^2}{E}\frac{dq}{du} = 3\left(1 + \frac{u^2}{E^2}\right)\left(1 - \frac{u}{E}\arctan\frac{E}{u}\right) - 1.$$

Hence, normal gravity at the equator ($\beta = 0$) and pole ($\beta = \pm 90°$) become:

$$\gamma_a = \frac{GM}{ab}\left[1 - m - \frac{me'}{6}\frac{q'(b)}{q(b)}\right] \quad \text{and} \quad \gamma_b = \frac{GM}{a^2}\left[1 + \frac{me'}{3}\frac{q'(b)}{q(b)}\right], \tag{3.11b}$$

which two formulas can be combined to:

$$2\frac{\gamma_a}{a} + \frac{\gamma_b}{b} = 3\frac{GM}{a^2 b} - 2\omega^2. \tag{3.11c}$$

The rigorous form of Clairaut's theorem from 1738 relates the geometric and gravity flattening by the equation

$$\frac{a-b}{a} + \frac{\gamma_b - \gamma_a}{\gamma_a} = \frac{\omega^2 b}{\gamma_a}\left(1 + \frac{e'q'(b)}{2q(b)}\right), \tag{3.12}$$

which can be verified by inserting $b = a\sqrt{1-e^2}$ and (3.11b) into the left member of (3.12) and noting the definition of m above. Then one obtains:

$$\text{Left member} = 1 - \sqrt{1-e^2} + \frac{\gamma_b - \gamma_a}{\gamma_a} = \frac{\gamma_b - \gamma_a\sqrt{1-e^2}}{\gamma_a} = \text{Right member}.$$

Somagliana (1929) derived the following elegant closed formula for the normal gravity on the ellipsoid:

$$\gamma_0 = \frac{a\gamma_b\sin^2\beta + b\gamma_a\cos^2\beta}{\sqrt{a^2\sin^2\beta + b^2\cos^2\beta}} = \frac{a\gamma_a\cos^2\varphi + b\gamma_b\sin^2\varphi}{\sqrt{a^2 cos^2\varphi + b^2\sin^2\varphi}}, \tag{3.13a}$$

which can also be written (Moritz 2000):

$$\gamma_0 = \gamma_a\frac{1 + k\sin^2\varphi}{\sqrt{1 - e^2\sin^2\varphi}}; \quad k = \frac{b\gamma_b}{a\gamma_a} - 1. \tag{3.13b}$$

It is obvious that normal gravity on the ellipsoid depends only on the four parameters a, b, γ_a and γ_b.

One may also show that γ changes approximately with elevation h by:

$$\gamma - \gamma_0 = -\frac{2h}{a}\gamma_a\left[1 + f + m + \left(\frac{5}{2}m - 3f\right)\sin^2\varphi\right] + \frac{3\gamma_a}{a^2}h^2, \tag{3.14a}$$

or, with constants taken from GRS1980, as:

$$\gamma - \gamma_0 = -\gamma_a\left[\left(a_1 + a_2\sin^2\varphi\right)h + a_3h^2\right], \tag{3.14b}$$

where $a_1 = 3.157 \times 10^{-7}$, $a_2 = -4.496 \times 10^{-10}$ and $a_3 = -7.374 \times 10^{-14}$.

For a rough estimation of the vertical gradient of normal gravity, one can start from:

$$\gamma \approx \frac{GM}{r^2},$$

which yields:

$$\frac{\partial\gamma}{\partial h} \approx \frac{\partial\gamma}{\partial r} \approx -2\frac{\gamma}{r} \quad \text{and} \quad \frac{\partial^2\gamma}{\partial h^2} \approx \frac{\partial^2\gamma}{\partial r^2} \approx 6\frac{\gamma}{r^2}. \tag{3.15}$$

According to GRS80 the geometric and gravimetric flattenings are $f^{-1} = 298.257222$ and $f^{*-1} = \gamma_e/(\gamma_p - \gamma_e) = 188.5924$, respectively.

3.2.3 The Disturbing Potential, Geoid and Gravity Anomaly

The geoid is an important reference surface in geodesy and geophysics. C. F. Gauss (1777–1855) introduced it as a model for the physical figure of the Earth, and he also *defined* it as the equipotential surface of the gravity field that coincides with mean sea level over the oceans (and it continues as an invisible level surface through the continents; Gauss 1828). From a practical point of view, the definition must also treat *the permanent tidal potential* in one way or another. In 1983, IAG adopted the *zero-geoid* for geoid determinations, implying that tidal attraction is removed from gravity data, while the permanent tidal deformation is retained. See also Ekman (1989). (This definition agrees with the standard definition of orthometric and normal heights.)

Definition 1 The geoid is the equipotential surface (level surface) of the Earth's gravity field that most closely coincides with the undisturbed mean sea level (and its continuation through the continents).

Disturbances are caused by ocean tides, streams, winds, variations in salinity and temperature, etc, of the order of $\pm$ 2 m.

Definition 2 The reference ellipsoid is an ellipsoid whose centre coincides with the gravity centre of the Earth (geo centre) and its dimensions semi-major and semi-minor axes are those that most closely agree with the geoid in a global sense. Moreover, the potential of the reference ellipsoid (the normal potential) on the ellipsoid equals the potential of the Earth (the geopotential) at the geoid.

In Sect. 7.4, we briefly discuss how to determine the potential of the geoid.

As both the geopotential (W) and the normal potential (U) are numerically large quantities, it is convenient to introduce the *disturbing potential*, $T = W - U$, for numerical computations. As both the geopotential and the reference ellipsoid (but not T) include the rotational potential Ω, it is obvious that (only) the disturbing potential is harmonic outside the Earth's mass (when neglecting the atmosphere).

As stated above, gravity is the gradient of the geopotential along the plumb line (with opposite sign), and normal gravity is the gradient of the normal potential along the normal to the reference ellipsoid (with opposite sign), which can be formulated:

$$g = -\frac{\partial W}{\partial n'} \approx -\frac{\partial W}{\partial h} \approx -\frac{\partial W}{\partial r} \quad \text{(gravity)} \tag{3.16}$$

and

$$\gamma = -\frac{\partial U}{\partial n} = -\frac{\partial U}{\partial h} \approx -\frac{\partial U}{\partial r} \quad \text{(normal gravity)}. \tag{3.17}$$

Here n' is the tangent of the plumb line, n is the normal to the reference ellipsoid through the observation point, r is the radius from the geocentre and h is the height along the normal to the ellipsoid. From these quantities, one obtains the residual quantities

$$\delta g = g - \gamma \approx -\frac{\partial T}{\partial r} = \text{gravity disturbance} \tag{3.18}$$

and

$$\Delta g = g_g - \gamma_0 = \text{gravity anomaly}, \tag{3.19}$$

where g_g and γ_0 denote gravity at the geoid and normal gravity on the reference ellipsoid, respectively. Traditionally, only the gravity anomaly could be well-determined as the gravity disturbance requires that the geodetic/ellipsoidal height of the gravity observation is known (which was not the case prior to the advent of GNSS in the positioning of the gravity observation points).

Definition 3 The geoid height (N) is the normal distance from the reference ellipsoid to the geoid (see Fig. 1.4.)

By making a Taylor expansion of the normal potential at the reference ellipsoid along the ellipsoidal normal, one obtains for the disturbing potential on the geoid (denoted by subscript g):

$$T_g = W_g - U_g = U_0 - (U_0 + N\frac{\partial U}{\partial h} + \cdots) = N\gamma_0 + \cdots, \tag{3.20}$$

where U_0 is the normal potential on the reference ellipsoid. Hence, one obtains the approximate *Bruns' formula* (Bruns 1878)

$$N = \frac{T_g}{\gamma_0}, \tag{3.21}$$

where the error is less than 0.02%.

By a similar Taylor series for γ:

$$\frac{\partial T}{\partial h} \approx -g + \gamma \approx -g + \gamma_0 + N\frac{\partial \gamma}{\partial h}, \tag{3.22}$$

one readily arrives at "*the boundary condition*" or "*fundamental equation*" of physical geodesy:

$$\Delta g = -\frac{\partial T}{\partial h} + \frac{\partial \gamma}{\partial h}\frac{T}{\gamma}. \tag{3.23a}$$

This relation holds on the geoid. Its *spherical approximation,* derived next, is most important in physical geodesy. Starting from the approximations $\partial h \approx \partial r$ and $\gamma \approx GM/r^2$, where GM = gravitational constant times mass of the Earth, one obtains $\frac{\partial \gamma}{\partial r} \approx -2\gamma/r$, which expression inserted into Eq. (3.23a) yields the spherical approximation of the boundary condition:

$$\Delta g = -\frac{\partial T}{\partial r} - 2\frac{T}{r} \tag{3.23b}$$

This first-order differential equation in T is *the fundamental equation of physical geodesy*. Its solution was first shown by Stokes (1849). One solution is presented in Sect. 3.3.2.

Note that in Eq. (3.19), we introduced the classical gravity anomaly on the geoid. In modern physical geodesy (see Chap. 4), there is no such limitation, but the definitions of the gravity anomaly as well as the boundary condition (Eqs. 3.23a and 3.23b) apply to any point in space.

Exercise 3.3 Show that, under spherical approximation, $r\Delta g$ is harmonic in the space outside masses. Hint: Consider that

$$r\partial T/\partial r = x\partial T/\partial x + y\partial T/\partial y + z\partial T/\partial z.$$

The solution is given in Appendix 2.

3.2.4 Harmonic Expansion of the Gravity Field

Spherical harmonics were introduced in Sect. 2.5. The disturbing potential can be expressed as a series in spherical harmonics on and outside a sphere of radius R (= the mean sea-level radius) by:

$$T = T(r,\Omega) = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n} T_{nm}Y_{nm}(\Omega), \tag{3.24a}$$

or

$$T = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1} T_n(\Omega), \tag{3.24b}$$

where (r,Ω) is the 3D-position with r being the geocentric radius, $\Omega = (\theta,\lambda)$ and $T_n(\Omega)$ are Laplace harmonics. Inserting (3.24a) into the boundary condition (3.23b), one obtains a harmonic series for the gravity anomaly:

$$\begin{aligned}\Delta g = \Delta g(r,\Omega) &= \sum_{n=0}^{\infty}\frac{n-1}{R}\left(\frac{R}{r}\right)^{n+2}\sum_{m=-n}^{n} T_{nm}Y_{nm}(\Omega)\\ &= \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+2}\sum_{m=-n}^{n}\Delta g_{nm}Y_{nm}(\Omega)\end{aligned} \tag{3.25a}$$

or

$$\Delta g = \sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+2}\Delta g_n(\Omega). \tag{3.25b}$$

The harmonics for T and Δg on the sphere are thus related by:

$$\Delta g_{nm} = \frac{n-1}{R}T_{nm} \Leftrightarrow \Delta g_n = \frac{n-1}{R}T_n, \tag{3.26a}$$

where

$$\begin{pmatrix} T_{nm} \\ \Delta g_{nm}\end{pmatrix} = \frac{1}{4\pi}\iint_{\sigma}\begin{pmatrix} T \\ \Delta g\end{pmatrix}Y_{nm}d\sigma \tag{3.26b}$$

and

$$\begin{pmatrix} T_n \\ \Delta g_n\end{pmatrix} = \begin{pmatrix} T_n(\Omega) \\ \Delta g_n(\Omega)\end{pmatrix} = \frac{2n+1}{4\pi}\iint_{\sigma}\begin{pmatrix} T \\ \Delta g\end{pmatrix}P_n(\cos\psi)d\sigma. \tag{3.26c}$$

It can be shown that Eqs. (3.24a), (3.24b), (3.25a) and (3.25b) are the solutions to Laplace's equations in the exterior of the sphere for T and $r\Delta g$, respectively. See, e.g. Heiskanen and Moritz (1967, Sects. 1.8 and 1.9).

From Eq. (3.25a), one can also derive the radial derivative of the gravity anomaly as:

$$\frac{\partial \Delta g}{\partial r} = -\sum_{n=0}^{\infty} \frac{(n-1)(n+2)}{R^2} \left(\frac{R}{r}\right)^{n+3} T_n(\Omega). \tag{3.27}$$

Comparing the series for T, Δg and $\partial \Delta g/\partial r$, one can see that the relative sensitivity (the power) within each parameter is pushed from 1 to (n − 1) and finally to $(n - 1)(n + 2)$. This is also seen in the power spectra (degree variances):

$$\begin{bmatrix} c_n^2(T) \\ c_n^2(\Delta g) \\ c_n^2(\partial \Delta g/\partial r) \end{bmatrix} = \frac{1}{4\pi} \iint_{\sigma} \begin{Bmatrix} T_n^2 \\ \Delta g_n^2 \\ (\partial \Delta g/\partial r)_n^2 \end{Bmatrix} d\sigma = \begin{bmatrix} 1 \\ (n-1)^2/R^2 \\ (n-1)^2(n+1)^2/R^4 \end{bmatrix} \sum_{m=-n}^{n} T_{nm}^2 \tag{3.28}$$

Hence, by observing gravity anomalies, and, in particular, its gradient, one can determine more details of the gravity field then from the geopotential.

Already in 1963 (a few years after the launch of the first artificial satellite), W. Kaula predicted that the geopotential harmonics attenuate by ("Kaula's rule"; Kaula 1966)

$$O(C_{nm}, S_{nm}) \approx 10^{-5}/n^2, \tag{3.29}$$

a formula that later has proved to be fairly accurate, and, even today, it is used for comparison with the power spectra of Earth gravitational models and their error estimates (see Chap. 4). In the determination of pure satellite-based EGMs, Eq. (3.29) is commonly used to stabilize the solution to higher degrees than otherwise would be possible, implying that many (old) EGMs are biased towards Kaula's rule.

3.3 Integral Formulas in Physical Geodesy

Several integral formulas on the sphere are of fundamental importance in classical physical geodesy. These are Poisson's, Stokes', Hotine's and Vening Meinesz' formulas as well the integral formula for gravity gradients. All these equations, presented below, integrate gravity anomalies on the sphere.

3.3.1 Poisson's Integral

Consider the series expansion (3.24b) for the disturbing potential

$$T_P = \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} T_n(\Omega),\ r_P \geq R, \tag{3.30}$$

where $T_n(\Omega)$ are the Laplace harmonics of T on the sphere of radius R given by Eq. (3.26c), which inserted into (3.30) and, after inter-changing the order of integral and summation, yields Poisson's formula in the exterior case ($r_P \geq R$):

$$T_P = \frac{1}{4\pi} \iint_{\sigma} TP(r_P, \psi) d\sigma, \tag{3.31a}$$

where

$$P(r_P, \psi) = \sum_{n=0}^{\infty} (2n+1) \left(\frac{R}{r_P}\right)^{n+1} P_n(\cos\psi), \tag{3.31b}$$

or, in a closed form (see Eq. 2.39)

$$P(r_P, \psi) = \frac{R\left(r_P^2 - R^2\right)}{l_P^3}; \quad \text{where } l_P = \sqrt{r_P^2 + R^2 - 2r_P R\cos\psi} \tag{3.31c}$$

is Poisson's kernel function in the external case of the sphere. (Frequently the harmonics of degrees zero and one are excluded from the kernel.)

As $r\Delta g$ is harmonic (under spherical approximation), it can also be determined by Eq. (3.31a) in the exterior space from surface values on the sphere.

3.3.2 Stokes' Formula

Stokes' formula (Stokes 1849) is the solution to the differential equation Eq. (3.23b) in the exterior case. By expressing T_n in Δg_n (see Eq. 3.26a) and inserting it into Eq. (3.30), one obtains a new series for T:

$$T_P = \sum_{n=2}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} \frac{R}{n-1} \Delta g_n(\Omega_P),\ r_P \geq R, \tag{3.32}$$

and, by considering (3.26c), one obtains *Stokes' extended formula* and kernel function after some rearrangements of terms (Pizzetti 1911):

$$T_P = \frac{R}{4\pi} \iint_{\sigma} S(r_P, \psi) \Delta g d\sigma, \quad r_P \geq R \tag{3.33a}$$

where

$$S(r_P, \psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r_P} \right)^{n+1} P_n(cos\psi), \tag{3.33b}$$

or, in a closed form,

$$S(r_P, \psi) = 2\frac{R}{l_P} + \frac{R}{r_P} - 3\frac{Rl_P}{r_P^2} - \frac{R^2}{r_P^2} \cos\,\psi \left(5 + 3\ln \frac{r_P - R\,\cos\,\psi + l_P}{2r_P} \right). \tag{3.33c}$$

Note that Eqs. (3.32)–(3.33c) lack terms of degrees zero and one. For $r_P = R$, Eqs. (3.33a) and (3.33b) become the traditional Stokes' formula and function. The latter can be expressed:

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi). \tag{3.34a}$$

or, in a closed form:

$$S(\psi) = 1 + \sin^{-1}\omega - 5\cos(2\omega) - 6\sin\omega - 3\cos(2\omega)\ln\left(\sin\omega + \sin^2\omega\right); \quad \omega = \psi/2. \tag{3.34b}$$

If Eq. (3.33a) with $r_P = R$ is inserted into Bruns' formula Eq. (3.21), one finally obtains *Stokes' original formula* for the geoid height:

$$N = \frac{R}{4\pi\gamma_0} \iint_{\sigma} S(\psi) \Delta g d\sigma. \tag{3.35}$$

This is a most important formula in physical geodesy. For a derivation in the space domain, see Heiskanen and Moritz (1967, Sects. 2–16). The integral contains no harmonics of degrees zero and one. As will be seen later, it can be applied in a variety of ways for geoid determination. This is particularly the case when combining it with an Earth gravitational model (see Chaps. 4 and 6).

Exercise 3.4 Consider Eqs. (3.24b) and the b.v.p. (3.23b) and Eq. (3.26c). Prove Stokes' formula. Hint: Assume that T can be expressed by the series $T = \sum_{n=2}^{\infty} \Omega_n \Delta g_n$ for some coefficients Ω_n.

A solution is given in Appendix 2.

Note: The solution (3.35) *holds only if* (a) Eq. (3.23b) *holds strictly [⇒ physical part of the direct ellipsoidal effect],* (b) *the integration takes place on a sphere* [⇒ *the dwc effect and the geometric part of the direct ellipsoidal effect] and c) there are no masses outside the sphere [⇒ the direct topographic and atmospheric effects].* (See Sects. 3.4 and 5.4).

Stokes' formula is the solution of *a free boundary value problem* in the sense that the boundary, the geoid, is not known.

3.3.3 Hotine's Formula

Differentiating Eq. (3.24b) w.r.t. r and changing sign, one obtains a harmonic series for the gravity disturbance in Laplace harmonics $T_n(\Omega)$:

$$\delta g = -\frac{\partial T}{\partial r} = \sum_{n=0}^{\infty} \frac{n+1}{R} \left(\frac{R}{r}\right)^{n+2} T_n(\Omega), \tag{3.36}$$

which yields the following relation between Laplace harmonics of the gravity disturbance and the disturbing potential:

$$\delta g_n = \frac{n+1}{R} T_n. \tag{3.37}$$

Continuing similarly to the derivation of Stokes' formula above, one finally arrives at *Hotine's* (*extended*) *formula* (Hotine 1969, p. 318)

$$T_P = \frac{R}{4\pi} \iint_{\sigma} H(r_P, \psi) \delta g d\sigma; \quad r_P \geq R, \tag{3.38}$$

where Hotine's (extended) function is:

$$H(r_P, \psi) = \sum_{n=0}^{\infty} \frac{2n+1}{n+1} s^{n+2} P_n(t), \quad \text{where } s = R/r_P \text{ and } t = \cos(\psi) \tag{3.39a}$$

or, in a closed form (see Appendix 1):

$$H(r_P, \psi) = \frac{2}{L} - \ln \frac{s - t + L}{1 - t}, \quad \text{where } L = \sqrt{1 - 2st + s^2}. \tag{3.39b}$$

Hotine's formula solves the Neumann problem or *the fixed bvp* for the disturbing potential on and outside the sphere that approximates the geoid. (A fixed bvp means that the boundary is known/fixed.) Note that, in contrast to Stokes' function, Hotine's function includes terms of degrees zero and one. Inserting (3.38) with $r_P = R$ into Bruns' formula, the geoid height is obtained.

3.3.4 *Vening Meinesz' Integrals*

The plumb line deviates from the normal to the reference ellipsoid, and this angle Θ is called the deflection of the seen in Fig. 3.2, the deflection in an arbitrary direction is related to the infinitesimal change of the geoid height (dN) vs. an infinitesimal distance ds, i.e.

$$\Theta = -\frac{dN}{ds}, \tag{3.40}$$

where the minus is a sign convention. Traditionally, Eq. (3.40) was used as a finite approximation in geodetic triangulation networks to determine geoid height differences by astronomic deflections and trigonometric height differences (so-called astrogeodetic determination of the geoid; see, e.g. Heiskanen and Moritz 1967, Sects. 5–7). In particular, in the north-south and east-west directions, the deflection components become:

$$\xi = -\frac{\partial N}{\partial x} = -\frac{\partial N}{R\partial\varphi} \tag{3.41a}$$

and

$$\eta = -\frac{\partial N}{\partial y} = -\frac{\partial N}{R\cos\varphi\,\partial\lambda}, \tag{3.41b}$$

respectively, and the total deflection is:

$$\Theta = \sqrt{\xi^2 + \eta^2}. \tag{3.41c}$$

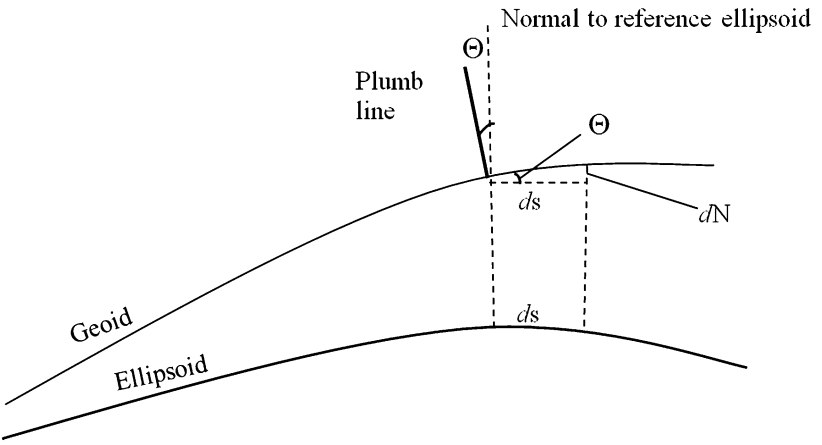

Fig. 3.2 Deflection of the vertical versus ellipsoidal normal

Applying the chain-rule one also obtains:

$$\xi = -\frac{\partial N}{\partial \psi}\frac{\partial \psi}{\partial x} \quad \text{and} \quad \eta = -\frac{\partial N}{\partial \psi}\frac{\partial \psi}{\partial y}, \tag{3.42}$$

and, by inserting Stokes' formula (3.35) into (3.42) and considering the spherical trigonometric relations:

$$\frac{\partial \psi}{\partial x} = -\frac{\cos\alpha}{R} \quad \text{and} \quad \frac{\partial \psi}{\partial y} = -\frac{\sin\alpha}{R}, \tag{3.43}$$

one finally arrives at *Vening Meinesz' formulas* (Vening Meinesz 1928):

$$\begin{Bmatrix} \xi \\ \eta \end{Bmatrix} = \frac{1}{4\pi\gamma_0}\iint_{\sigma} \begin{Bmatrix} \cos\alpha \\ \sin\alpha \end{Bmatrix} \frac{dS(\psi)}{d\psi}\Delta g d\sigma \tag{3.44}$$

Here α is the azimuth from the computation point to the integration point.

3.3.5 The Vertical Gradient of Gravity

The vertical gradient of gravity is sometimes of interest, for example, for a precise downward continuation of the gravity observation on the Earth's surface to the geoid. The forward problem (to determine the gradient outside a sphere from gravity anomalies on the sphere) can be obtained in the space domain from Poisson's integral (3.31a) for the harmonic function $r\Delta g$ as:

$$\left(\frac{\partial(r\Delta g)}{\partial r}\right)_P = \frac{R}{4\pi}\iint_{\sigma} M(r_P,\psi)\Delta g d\sigma; \quad r_P > R \tag{3.45a}$$

with the kernel function

$$M(r_P,\psi) = \frac{\partial}{\partial r_P}P(r_P,\psi) = \frac{R\left(r_P^2+R^2\right)}{r_P l_P^3} - \frac{R\left(r_P^2-R^2\right)^2}{r_P l_P^5}, \tag{3.45b}$$

or

$$\left(\frac{\partial(\Delta g)}{\partial r}\right)_P = -\frac{\Delta g_P}{r_P} + \frac{R}{4\pi}\iint_{\sigma} M(r_P,\psi)\Delta g d\sigma. \tag{3.45c}$$

In the spectral domain, one obtains from the harmonic series (3.25b):

$$\left(\frac{\partial \Delta g}{\partial r}\right)_P = -\sum_{n=0}^{\infty} \frac{n+2}{R}\left(\frac{R}{r_P}\right)^{n+3} \Delta g_n(\Omega); \quad r_P \geq R. \tag{3.46}$$

On the sphere with $r_P = R$, Eq. (3.45b) becomes (Heiskanen and Moritz 1967, Sects. 1–18)

$$M(\psi) = \frac{2R^2}{l_0^3}; \quad l_0 = 2R \sin\frac{\psi}{2}, \tag{3.47}$$

and (3.45c) takes the form

$$\left(\frac{\partial \Delta g}{\partial r}\right)_P = \frac{R^2}{2\pi}\iint_{\sigma} \frac{\Delta g}{l_0^3} d\sigma = -\frac{\Delta g_P}{R} + \frac{R^2}{2\pi}\iint_{\sigma} \frac{\Delta g - \Delta g_P}{l_0^3} d\sigma; \quad (r_P = R), \tag{3.48}$$

where the last form is convenient for numerical integration. The function $M(\psi)$ is very local and attenuates quickly towards zero for increasing ψ, which implies that it is usually sufficient to integrate in a small area/cap around the computation point.

3.3.6 *The Inverse Vening Meinesz Formula*

The following solution for the inverse Vening Meinesz formula is inspired by a solution by Hwang (1998), but it is slightly different.

According to Eqs. (3.41a), (3.41b) and the notation introduced in Eq. (2.129b), the deflection components of the vertical are related with the geoid height by the compact formula

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = -\nabla N_Q = -\frac{1}{R}\nabla^* N_Q. \tag{3.49}$$

A tentative solution for the *inverse Vening Meinez formula* can therefore be expressed:

$$\Delta g_P = \frac{\gamma_0}{4\pi}\iint_{\sigma} \left[\nabla^* F_1\left(\psi_{PQ}\right)\right]^T \bullet \nabla N_Q d\sigma_Q, \tag{3.50a}$$

where the kernel function is of the form

$$F_1(\psi) = \sum_{n=2}^{\infty} F_{1,n} P_n(\cos\psi), \tag{3.50b}$$

and its coefficients remain to be determined. To do so, we express the geoid height as a harmonic series with coefficients N_{nm}, i.e.

$$N_Q = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} N_{nm} Y_{nm}(Q), \tag{3.51}$$

we use the addition theorem for spherical harmonics, Eq. (2.49), in Eq. (3.50b):

$$F_1(\psi_{PQ}) = \sum_{n=2}^{\infty} \frac{F_{1,n}}{2n+1} \sum_{m=-n}^{n} Y_{nm}(P) Y_{nm}(Q), \tag{3.52}$$

and we consider the inner product between the surface gradients of spherical harmonics from Eq. (2.130). Then Eq. (3.50a) becomes:

$$\Delta g_P = \frac{\gamma_0}{R} \sum_{n=2}^{\infty} F_{1,n} \frac{n(n+1)}{2n+1} \sum_{m=-n}^{n} N_{nm} Y_{nm}(P). \tag{3.53}$$

However, the gravity anomaly on the sphere can also be expressed [cf. (3.25a) with $r = R$]

$$\Delta g_P = \gamma_0 \sum_{n=2}^{\infty} \frac{n-1}{R} \sum_{m=-n}^{n} N_{nm} Y_{nm}(P), \tag{3.54}$$

and by comparing the spectra of (3.53) and (3.54) one arrives at

$$F_{1,n} = \frac{(2n+1)(n-1)}{n(n+1)}, \tag{3.55}$$

and therefore:

$$F_1(\psi) = \sum_{n=2}^{\infty} \frac{(2n+1)(n-1)}{n(n+1)} P_n(\cos\psi) = \sum_{n=2}^{\infty} \left[2 - \frac{1}{n} - \frac{1}{1-n}\right] P_n(\cos\psi). \tag{3.56}$$

As shown by Hwang (1998), the closed form of this kernel is (see also Appendix 1 for $s = 1$)

$$F_1(\psi) = \frac{1}{\upsilon} + \ln\left(\frac{\upsilon^3}{1+\upsilon}\right) \tag{3.57}$$

and its derivative w.r.t. ψ can be expressed:

$$F_1{}'(\psi) = \left(\frac{3}{\upsilon} - \frac{1}{\upsilon^2} - \frac{1}{1+\upsilon}\right)\frac{\cos(\psi/2)}{2} \tag{3.58}$$

where $\upsilon = \sin(\psi/2)$.

Hence, as

$$\nabla^* F_1(\psi_{PQ}) = -F_1{}'(\psi_{PQ})(\cos\alpha', \sin\alpha')^T \tag{3.59}$$

the closed form of Eq. (3.50a) becomes:

$$\Delta g_P = \frac{\gamma_0}{4\pi}\iint_\sigma F_1'(\psi_{PQ})\varepsilon_{QP}d\sigma_Q, \tag{3.60a}$$

where

$$\varepsilon_{QP} = \xi_Q\cos\alpha' + \eta_Q\sin\alpha' \tag{3.60b}$$

is the deflection at the running point Q with azimuth α' towards the computation point P.

3.3.7 The Geoid-from-Deflection Formula

By considering the Laplace spectral relation between the geoid height and gravity anomaly (Eq. 3.32)

$$N_n = \frac{R}{n-1}\frac{\Delta g_n}{\gamma_0}, \tag{3.61}$$

it follows from Eqs. (3.56) and (3.60a) that

$$N_P = \frac{R}{4\pi}\iint_\sigma F_2'(\psi_{PQ})\varepsilon_{QP}d\sigma_Q, \tag{3.62}$$

where in accord with Eq. (3.55),

$$F_2(\psi) = \sum_{n=2}^{\infty}\frac{(2n+1)}{n(n+1)}P_n(\cos\psi) = \sum_{n=2}^{\infty}\left[\frac{1}{n} + \frac{1}{n+1}\right]P_n(\cos\psi). \tag{3.63a}$$

Hwang (1998) derived the following closed forms of the kernel (see also Appendix 1 for $s = 1$):

$$F_2(\psi) = -1 - \frac{3}{2}\cos\psi - 2\ln[\sin(\psi/2)] \tag{3.63b}$$

and

$$F_2'(\psi) = -\cot\frac{\psi}{2} + \frac{3}{2}\sin\psi. \tag{3.64}$$

3.3.8 Gradiometry Formulas on the Sphere

The observables available in gradiometry can be described by the symmetric (gravity-) gradiometry matrix (the Eötvös tensor) $\mathbf{G}$, consisting of the following set of second-order derivatives (w.r.t. a local Cartesian coordinate system) of the geopotential:

$$\mathbf{G} = \begin{bmatrix} W_{xx} & W_{xy} & W_{xz} \\ W_{yx} & W_{yy} & W_{yz} \\ W_{zx} & W_{zy} & W_{zz} \end{bmatrix}. \tag{3.65a}$$

(If W is substituted by the gravitational potential, Eq. (3.65a) is named the Marussi tensor.) The matrix holds as well for the disturbing potential

$$\overline{\mathbf{G}} = \begin{bmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{bmatrix}, \tag{3.65b}$$

where $T_{xy} = T_{yx}$, $T_{xz} = T_{zx}$, and $T_{yz} = T_{zy}$. Also, by applying the Laplace equation in space, it holds that:

$$T_{xx} + T_{yy} + T_{zz} = 0, \tag{3.66}$$

and, therefore, there are only five out of nine independent observables in $\overline{\mathbf{G}}$.

Recalling the external type harmonic series for the disturbing potential

$$T = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} T_n; \quad r \geq R, \tag{3.67}$$

and remembering that the local z-axis is along the radius vector with magnitude r, it follows that:

$$-T_z = \delta g = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+2} \frac{n+1}{R} T_n, \tag{3.68a}$$

$$T_{zz} = -\delta g_z = \sum_{n=0}^{\infty} \frac{(n+1)(n+2)}{R^2} \left(\frac{R}{r}\right)^{n+3} T_n, \tag{3.68b}$$

$$T_{zx} = -\delta g_x = -\sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+2} \frac{n+1}{R} \frac{\partial T_n}{\partial x} = \frac{\gamma_0}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+2} (n+1)\xi_n, \tag{3.68c}$$

and

$$T_{zy} = -\delta g_y = -\sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+2} \frac{n+1}{R} \frac{\partial T_n}{\partial y} = \frac{\gamma_0}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+2} (n+1)\eta_n, \tag{3.68d}$$

where the last steps in (3.68c), (3.68d) follow from Eqs. (3.41a), (3.41b). All Laplace harmonics refer to the sphere of radius R. By setting $r = R$ in Eqs. (3.68c) and (3.68d), and equating each spectral harmonic, these equations can be inverted to:

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{R}{\gamma_0} \sum_{n=0}^{\infty} \frac{1}{n+1} \begin{pmatrix} T_{zx} \\ T_{zy} \end{pmatrix}_n, \tag{3.69}$$

or

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \frac{R}{4\pi\gamma_0} \iint_{\sigma} H(\psi) \begin{pmatrix} T_{zx} \\ T_{zy} \end{pmatrix} d\sigma, \tag{3.70}$$

which is Hotine's integral formula applied to the mixed zx- and zy-gradiometry components. As a result, we see that these components of the gradiometry matrix are closely related to the deflections of the vertical.

From Eqs. (3.69) and (3.70), one can derive the following relationships between Laplace harmonics on the sphere:

$$(T_{zz})_n = -\frac{n+2}{R} \delta g_n = -\frac{(n+2)(n+1)}{n-1} \frac{\Delta g_n}{R} = \frac{(n+2)(n+1)}{R^2} T_n, \tag{3.71}$$

and, by inverting each of these equations and summing up the harmonics, one obtains the harmonic series

$$\delta g = -R \sum_{n=0}^{\infty} \frac{(T_{zz})_n}{n+2}, \tag{3.72a}$$

$$\Delta g = -R\sum_{n=2}^{\infty}\frac{n-1}{(n+2)(n+1)}(T_{zz})_n, \tag{3.72b}$$

and

$$T = R^2\sum_{n=0}^{\infty}\frac{(T_{zz})_n}{(n+2)(n+1)}, \tag{3.72c}$$

which can also be expressed in the space domain by the spherical integrals

$$\delta g = -\frac{R}{4\pi}\iint_{\sigma} F_{1z}(\psi)T_{zz}d\sigma, \tag{3.73a}$$

$$\Delta g = -\frac{R}{4\pi}\iint_{\sigma} F_{2z}(\psi)T_{zz}d\sigma, \tag{3.73b}$$

and

$$T = \frac{R^2}{4\pi}\iint_{\sigma} F_{3z}(\psi)T_{zz}d\sigma, \tag{3.73c}$$

where, once again using the abbreviated notation $\upsilon = \sin(\psi/2)$,

$$F_{1z}(\psi) = \sum_{n=0}^{\infty}\frac{2n+1}{n+2}P_n(\cos\,\psi) = 1+\upsilon^{-1}-2\upsilon-t\,\ln\left(1+\upsilon^{-1}\right) \tag{3.74a}$$

$$F_{2z}(\psi) = \sum_{n=0}^{\infty}\frac{(2n+1)(n-1)}{(n+2)(n+1)}P_n(\cos\psi) = F_{1z}(\psi)-2F_{3z}(\psi) \tag{3.74b}$$

and

$$F_{3z}(\psi) = \sum_{n=0}^{\infty}\frac{2n+1}{(n+2)(n+1)}P_n(\cos\psi) = 6\upsilon-3+(3t-1)\quad\ln\left(\frac{1+\upsilon}{2}\right). \tag{3.74c}$$

These equations follow from Appendix 1 with $s = 1$.

Next we consider the horizontal components of gradiometry for gravity and geoid determination. Considering the derivation of the inverse Vening Meinesz formula in Sect. 3.3.6 and that $\nabla^* T_z = \begin{pmatrix} T_{zx} & T_{zy}\end{pmatrix}^T$, one may start from the integral equation

$$\delta g_P = -\frac{1}{4\pi}\iint_{\sigma}[\nabla^* K(\psi)]^T \bullet \nabla^* T_z d\sigma = \frac{R}{4\pi}\iint_{\sigma} K'(\psi) G_{QP} d\sigma, \tag{3.75a}$$

where

$$G_{QP} = (T_{zx})_Q \cos\alpha' + (T_{zy})_Q \sin\alpha', \tag{3.75b}$$

and the problem is to solve for the coefficients of the kernel function

$$K(\psi) = \sum_{n=0}^{\infty} K_n P_n(\cos\psi). \tag{3.75c}$$

Considering Eqs. (3.50a) and (3.53), one can see that the first equation in formula (3.75a) can be expressed by the Laplace series

$$\delta g_P = -\sum_{n=0}^{\infty} K_n \frac{n(n+1)}{2n+1}(T_z)_n. \tag{3.76a}$$

However, as there is also the simple relation among the Laplace harmonics for δg and T_z;

$$\delta g_n = -(T_z)_n, \tag{3.76b}$$

the solution for K_n becomes

$$K_n = \frac{2n+1}{n(n+1)} \quad \text{for } n \geq 1. \tag{3.76c}$$

If the summation starts from $n = 2$, $K(\psi)$ equals the previously derived kernel $F_2(\psi)$ of Eqs. (3.63a) and (3.63b) with opposite sign, and $F_2'(\psi)$ was given by Eq. (3.64). Hence, the integral formula for the gravity disturbance becomes:

$$\delta g_P = \frac{R}{4\pi}\iint_{\sigma} F_2'(\psi_{PQ}) G_{QP} d\sigma_Q. \tag{3.77}$$

Moreover, noting the following relationships between some Laplace harmonics

$$\Delta g_n = \frac{n-1}{n+1}\delta g_n \quad \text{and} \quad N_n = \frac{R}{n+1}\frac{\delta g_n}{\gamma_0},$$

it follows from Eq. (3.76c) that:

$$\Delta g_P = \frac{R}{4\pi} \iint_{\sigma} H_2'(\psi_{PQ}) G_{QP} d\sigma_Q, \tag{3.78a}$$

where, $H_2'(\psi_{PQ})$ is the derivative of $H_2(\psi_{PQ})$:

$$H_2(\psi) = \sum_{n=2} \frac{(2n+1)(n-1)}{n(n+1)^2} P_n(\cos\psi) \tag{3.78b}$$

w.r.t ψ, and

$$N_P = \frac{R^2}{4\pi\gamma_0} \iint_{\sigma} H_3'(\psi_{PQ}) G_{PQ} d\sigma_Q, \tag{3.78c}$$

where, again, H_3' is the derivative of:

$$H_3(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n(n+1)^2} P_n(\cos\,\psi). \tag{3.78d}$$

Closed-form solutions are tricky to derive for Eqs. (3.78b) and (3.78d), but for numerical applications the series can be tabulated. However, applying Eq. (8.121), closed forms of H_2' and H_3' can be obtained. (see the derivation of Eq. 8.122b.)

3.4 Practical Considerations (DITE, DWC, SITE, PITE)

Traditionally, the observed gravity is continued downward to sea level by a free-air correction, and the topographic signal is removed by applying the Bouguer correction and forming the Bouguer anomaly. However, the Bouguer anomaly is usually large, and therefore the topographic effect is usually reduced by adding a compensation attraction (see Sect. 3.4.3). All these corrections are applied as a *regularization*, implying that the reduced gravity anomaly satisfies the spherical approximation of the boundary condition of physical geodesy, whose solution is Stokes' formula.

3.4.1 The Free-Air Correction

Equation (3.19) shows that the traditional gravity anomaly is located on the geoid, which implies that the observed gravity value at the Earth's surface must be continued downward to the geoid level. This is handled by adding the free-air correction

$$F = -\frac{\partial \gamma}{\partial r} H \approx 0.3086H \,[\text{mGal for } H \text{ in m}], \tag{3.79}$$

which is approximated as the vertical gradient of gravity, and the radial derivative of the normal gravity stems from the approximation

$$\gamma \approx GM/r^2 \Rightarrow \frac{\partial \gamma}{\partial r} = -\frac{2\gamma}{r}, \tag{3.80}$$

Hence, the free-air gravity anomaly becomes:

$$\Delta g = g + F - \gamma_0. \tag{3.81}$$

3.4.2 The Bouguer Correction

The free-air gravity anomaly cannot be directly used in the integral formulas above, because they are not valid if there are masses external to the sphere of integration. Hence, the attraction of the topographic masses (A^T) should be removed. The resulting gravity anomaly is the simple or refined Bouguer gravity anomaly:

$$\Delta g^B = g + F - \gamma_0 - B, \tag{3.82}$$

implying that A^T is approximated by B the attraction of the Bouguer plate, an infinite plate of thickness H. If we consider the standard topographic density $\rho = 2.67$ g/cm^3, the Bouguer correction becomes (Heiskanen and Moritz 1967, p. 131):

$$B = -2\pi G\rho H = -0.1119H \,[\text{mGal for } H \text{ in m}] \tag{3.83}$$

where G is the gravitational constant. As a result the simple Bouguer anomaly can be written

$$\Delta g^B = g - \gamma_0 + 0.1967H \,[\text{mGal for } H \text{ in m}]. \tag{3.84}$$

The refined Bouguer anomaly implies that the topographic correction is improved by also adding an additional correction for the attraction of the terrain. The goal is to remove the total topographic attraction A^T (at the sphere of integration). In practice, the Stokesian integration was frequently carried out as a plane approximation, where the area surrounding the computation point was divided into sectors and compartments, and the topographic attraction of such elements could easily be computed based on the attraction of a homogeneous cylinder (see Heiskanen and Moritz 1967, Chap. 3), and for a refined solution a small correction could also be added for the spherical shape of the Earth.

3.4.3 The Direct Topographic Effect (DITE)

However, the Bouguer gravity anomaly is not a suitable gravity anomaly to apply in the integral formulas, because this anomaly is usually systematically large and has several disadvantages for applications in physical geodesy (see following sections). Instead one first *compensates* for the topographic attraction, implying that a model compensation attraction A^c is added, yielding *the topography compensated gravity anomaly*

$$\Delta g^c = \Delta g^B + A^c. \tag{3.85}$$

There are numerous compensation models, having in common that they compensate for the topographic attraction on or below sea level, such that (theoretically) there are no remaining masses above this level. Among a variety of models, we here just mention the two most commonly used ones: Helmert condensation-layering compensation and isostatic compensation. Helmert's second method of condensation method means that the topography is condensed as a surface layer at sea level with surface density ρH being the density times the height of the topography. A more advanced compensation model is provided by a topographic/isostatic model (see Sect. 8.4), and it usually provides the best compensation for the topographic attraction. However, while Helmert's compensation is easy to apply, the numerical effort to provide the isostatic compensation attraction is numerically more demanding.

The application of the topographic and compensation corrections $(-A^T + A^c)$ is called *the Direct Topographic Effect* (*DITE*) on gravity. It implies that the condition for applying the integral formulas with no masses permitted external to the sphere (if the atmosphere is neglected) is satisfied.

Traditionally, the application of the DITE has been called *regularization*.

3.4.4 The SITE, Co-geoid and the PITE

As the DITE implies a reduction for the attraction of the topography, the reduced gravity anomaly Δg^c on the geoid (approximated by the sphere) also needs a correction called the Secondary Indirect Topographic Effect (*SITE*) before Stokes' integration, and, after Stokes' integration, the resulting co-geoid height needs a correction by the Primary Indirect Topographic Effect (PITE), which is the effect corresponding to the restoration of the topographic masses. Hence, the geoid height is finally given by:

$$N = \frac{R}{4\pi\gamma_0} \iint_{\sigma} S(\psi)(\Delta g^c + SITE)d\sigma + PITE, \tag{3.86a}$$

where the integral yields the co-geoid height,

$$SITE = \delta N^c \frac{\partial \gamma}{\partial r} \approx -2\frac{\gamma_0}{R}\delta N^c \tag{3.86b}$$

and

$$PITE = \delta N^c = \frac{V_g^T - V_g^c}{\gamma_0}. \tag{3.86c}$$

The SITE is motivated to augment Δg^c from the geoid to the co-geoid for the Stokes integration. In Eq. (3.86c), V^T and V^c are the topographic and compensation potentials, respectively. Similarly, PITEs can be applied also in the other integral formulas.

In principal, all compensation models result in the same geoid heights. However, in practice there are differences due to interpolation and numerical integration of the discrete gravity observations.

If no compensation model is applied, the (Bouguer) gravity anomaly under Stokes' integral is large, and the *PITE* is enormous, even infinite for a flat Earth approximation. Although there are many gravity reduction models in the literature, the most common ones are Helmert's method of condensation, where the topography is modelled by a surface density, e.g. in Helmert's 2nd compensation model by ρH, where ρ and H are topographic density and height, respectively, yielding the compensation potential on the geoid:

$$V_g^c = GR \iint_\sigma \frac{\rho H}{l_0} d\sigma. \tag{3.87}$$

The application of this model is discussed in Martinec (1998) and Sjöberg (2000).

The classical definition of the gravity anomaly and its application to geoid determination as described above have some disadvantages that disqualify them for today's demands of accurate geoid determination. Although, most of the topographic effects (DITE, PITE and SITE) can be refined to meet the demands, the main problem lies in the downward continuation of gravity to the geoid to constitute the gravity anomaly on the geoid. As shown by Sjöberg (2014) the simple free-air reduction causes an error of the order of the geoid-to-quasigeoid separation, which increases from zero at sea level to some metres in the highest mountains. This problem is solved in the modern techniques as presented in Chaps. 4–6. As will be seen, also the motivation for adding the SITE differs from the reason discussed above.

3.5 Height Systems

Geometric heights (such as GNSS- derived heights and levelled heights) differ from *physical heights* related with the equipotential level surfaces of the Earth. National height systems are defined by some kind of physical height, while local systems used for surveying, construction and mapping sometimes are geometric. Physical heights, in contrast to geometric ones, are holonomic, implying that the difference between such heights at two points is independent on the path one takes between the two points. Here we are concerned with geopotential numbers as well as dynamic, orthometric and normal heights, which all belong to the category of physical heights. In many countries, one uses so-called normal-orthometric heights, also to be considered subsequently.

3.5.1 Geopotential Numbers

The geopotential number at point P is defined by:

$$C_P = W_g - W_P = \int_0^{H_P} g dh, \tag{3.88}$$

where H_P is the orthometric height (see Sect. 3.5.2), i.e. the height above the geoid along the plumb line to P. C_P is measured in geopotential units (g.p.u.), where 1 g.p.u. = 1 kGal m. As the geopotential number is not in units of metres, it is convenient to scale it. In this way the, dynamic height (adopted by the IAG in 1955) is obtained by:

$$H_P^{dyn} = \frac{C_P}{\gamma_{45}}, \tag{3.89}$$

where γ_{45} is normal gravity at latitude 45°. The dynamic correction, needed to correct the levelled height difference Δh_A^B between points A and B to become the dynamic height difference, is then given by:

$$DC_A^B = \frac{C_B - C_A}{\gamma_{45}} - \Delta h_A^B = \frac{1}{\gamma_{45}} \int_A^B (g - \gamma_{45}) dh \approx \frac{1}{\gamma_{45}} \sum_{i=A}^{B} (g_i - \gamma_{45}) dh_i, \tag{3.90}$$

where dh_i are individual height-difference measurements between the end points A and B. However, as normal gravity changes by as much as 2.6 Gal from the Equator

to mid-latitudes, also the dynamic correction is usually large. This drawback can be reduced by selecting a local reference γ instead of γ_{45} for a region, but then the regional type of dynamic heights must not be mixed with the IAG-adopted ones. The dynamic height is used in certain engineering projects, as it, in contrast to the levelled height, indicates in which direction a fluid flows.

3.5.2 Orthometric Heights

Many countries have adopted orthometric heights as their national height systems. Such a height is closely related to the geoid height, as their sum is the geodetic height, i.e. the height of the topographic surface above the reference ellipsoid (see Fig. 3.3). [In this context one disregards that the orthometric height actually is (very slightly) curved along the plumb line.] Applying the mean-value theorem of integral calculus to Eq. (3.88), one obtains the orthometric height H_P:

$$C_P = \bar{g} H_P \Rightarrow H_P = \frac{C_P}{\bar{g}}, \tag{3.91a}$$

where $\bar{g}$ is the mean value of gravity along the plumb line in the interval between the geoid and the surface point P:

$$\bar{g} = \frac{1}{H_P} \int_0^{H_P} g dh. \tag{3.91b}$$

The computation of an accurate $\bar{g}$ is complicated, as it requires that g is known all the way from the geoid to the surface. Frequently, it is approximated by the so-called Helmert-orthometric height, which uses the approximate mean-gravity value (Heiskanen and Moritz 1967, p. 167):

$$\bar{g}^H = g_P + 0.0424 H_P \,[\text{Gal for } H_P \text{ in km}]. \tag{3.92}$$

This approximation is based on so-called Prey reduction of gravity inside the topography with standard density 2.67 g/cm^3 according to the method of Poincare and Prey (Heiskanen and Moritz 1967, p. 163). The orthometric height may then need some iteration. For more accurate estimates of orthometric heights, see Sect. 7.2.2.

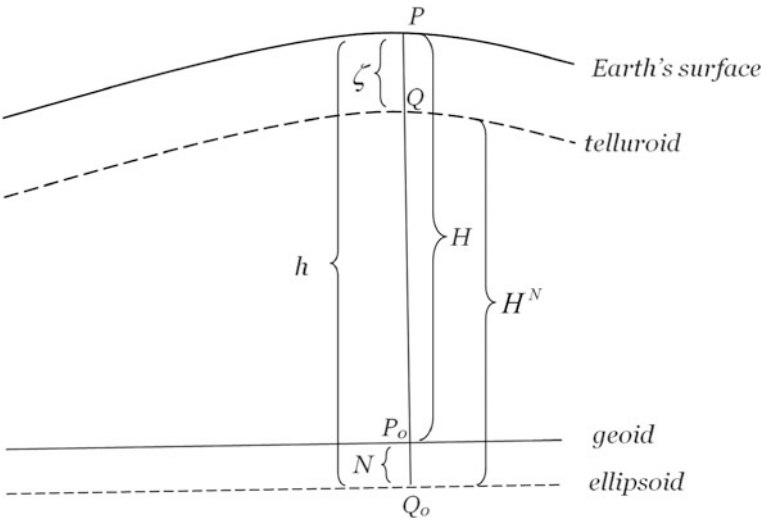

Fig. 3.3 Height systems (orthometric and normal heights)

As the sum of the geoid and orthometric heights (approximately) equals the geodetic height h_P (see Fig. 3.3), the orthometric height can also be determined geometrically by modern GNSS levelling:

$$H_P = h_p - N_P. \tag{3.93}$$

This technique requires that both the geoid height and the geometric height are known, the latter, e.g. from GNSS observations.

In ocean areas, H is usually denoted sea-surface topography, which is of basic interest in oceanography. It typically differs from the geoid height by $\pm$ 2 m due to salinity, currency, tides etc. The geodetic height (h) at sea is most successfully determined by satellite altimetry.

3.5.3 Normal Heights

The normal height belongs to the modern concepts of surface gravity anomaly, quasigeoid and height anomaly, introduced by M.S. Molodensky (Molodensky et al. 1962); see Chap. 4.

The normal height $\left(H_P^N\right)$ at a point P is the distance from the reference ellipsoid along the normal through P to the point Q on the *telluroid*. (Point Q is defined such that its normal potential U_Q equals W_P.) Then it holds that:

$$H_P^N = \frac{C_P}{\bar{\gamma}} \tag{3.94a}$$

where

$$\begin{aligned}\bar{\gamma} &= \frac{1}{H_P^N}\int_0^{H_P^N} \gamma dh = \frac{U_0 - U_Q}{H_P^N} \\ &= \frac{1}{H_P^N}\left[U_0 - \left(U_0 + (\frac{\partial U}{\partial h})_0 H_P^N + \frac{1}{2}\left(\frac{\partial^2 U}{\partial h^2}\right)_0 \left(H_P^2\right)^2 + \ldots\right)\right]\end{aligned} \tag{3.94b}$$

or, after simplifications:

$$\bar{\gamma} \approx \gamma_0 - \frac{\gamma_0}{R} H_P^N, \tag{3.94c}$$

which lends itself to iteration with Eq. (3.94a). Notice, that the normal height can be determined exactly for precise data with no assumption about the density of topography (in contrast to the orthometric height).

As the sum of the normal height and the height anomaly (ζ_P) is the geometric height (h_P) of the topography (see Fig. 3.3), i.e.

$$h = H + N = H^N + \zeta, \tag{3.95}$$

it follows that the normal height can also be determined by GNSS levelling as:

$$H_P^N = h_P - \zeta_P. \tag{3.96}$$

It is also possible to determine a precise orthometric height from the normal height and the so-called geoid-from-quasigeoid correction (*GQC*) (see Sect. 7.2), by:

$$H = H^N - GQC. \tag{3.97}$$

3.5.4 Normal-Orthometric Heights

The height systems in many countries are based on *normal-orthometric heights*, which is a compromise between orthometric and normal heights. It implies that the geopotential number is approximated by the known integral

$$\tilde{C}_P = \int_{P0}^{P} \gamma dh = U_{P0} - U_P, \tag{3.98}$$

where P is the surface point, and the height of the reference point P_0 is usually unknown, but can be assumed to be located close to the reference ellipsoid. The normal orthometric height is defined by:

$$H_P^{NO} = \frac{\tilde{C}_P}{\bar{\gamma}}, \tag{3.99}$$

where $\bar{\gamma}$ is the same mean normal gravity as used in the normal height. It is obvious that the normal-orthometric height is not based on gravity observations, but those data are replaced by normal gravity.

It is interesting to find out whether the normal-orthometric height can be corrected to the normal and/or orthometric heights. However, as the geopotential number here uses normal gravity instead of actual gravity observations, it appears rather impossible to add simple corrections for such improvements. Anyhow, some knowledge about such corrections can be achieved as follows.

– *The Corrections to Normal and Orthometric Heights*

By taking the difference between Eqs. (3.94a) and (3.99), one obtains the correction from normal-orthometric to normal height:

$$dH^{N-NO} = \frac{C - \tilde{C}}{\bar{\gamma}}, \tag{3.100}$$

which can also be expressed as:

$$dH^{N-NO} = \frac{W_g - W_P - U_{P0} + U_P}{\bar{\gamma}}. \tag{3.101}$$

Introducing

$$U_{P0} = U_0 + \delta U, \tag{3.102}$$

where U_0 and $\delta U = U_{P0} - U_0$ are the normal potential at the reference ellipsoid and its (unknown) correction to reach the reference potential used in the normal-orthometric height, respectively, and considering that $W_g = U_0$ and $W_P - U_P = T_P = \gamma_Q \zeta_P$ (Bruns' formula), one obtains:

$$dH^{N-NO} = -\frac{T_P + \delta U}{\bar{\gamma}} = -\frac{\gamma_Q}{\bar{\gamma}}\zeta - \frac{\delta U}{\bar{\gamma}} \approx -\zeta - \frac{\delta U}{\bar{\gamma}}. \tag{3.103}$$

In a similar way, one obtains the correction to the orthometric height:

$$dH^{O-NO} = \frac{C}{\bar{g}} - \frac{\tilde{C}}{\bar{\gamma}} = \frac{C-\tilde{C}}{\bar{\gamma}} + H\frac{\bar{\gamma}-\bar{g}}{\bar{\gamma}}, \tag{3.104}$$

which, in view of Eqs. (3.94a), (3.94b), (3.94c) and (3.103), becomes:

$$dH^{O-NO} \approx dH^{N-NO} - N + \zeta \approx -N - \frac{\delta U}{\bar{\gamma}}. \tag{3.105}$$

Here δU is constant, while $\bar{\gamma}$ changes with latitude (in γ_0) and elevation (see 3.94a, 3.94b and 3.94c). One way to fix δU would be to use a geoid or quasigeoid model as well as GNSS determined geodetic heights of one or (better) several points in the region to determine the orthometric or normal height(s) by GNSS levelling (Eqs. 3.93 and 3.95) with consideration of the corrective surface difference between the two types of heights. Then δU can be solved from either Eqs. (3.103) or (3.105).

Appendix 1: Closed-Form Kernels

Referring to Bois (1961), the following formulas can be derived for $0 \le s \le 1$ when using the notations $L(x) = \sqrt{(1-2xt+x^2)}$ and $L(s) = L$:

$$S_1 = \sum_{n=0}^{\infty} s^n P_n(t) = 1/L \tag{3.106}$$

$$\begin{aligned} S_2 &= \sum_{n=1}^{\infty} \frac{s^n}{n} P_n(t) = \int_0^s \sum_{n=1}^{\infty} x^{n-1} P_n(t) ds = \int_0^s \left(\frac{1}{xL(x)} - \frac{1}{x}\right) dx \\ &= -[\ln 2(1-xt+L)]_{x=0}^{s} = -\ln(1-st+L) + \ln 2 \end{aligned} \tag{3.107}$$

$$\begin{aligned} S_3 &= \sum_{n=0}^{\infty} \frac{s^{n+1}}{n+1} P_n(t) = \int_0^s \sum_{n=0}^{\infty} x^n P_n(t) dx = \int_0^s \frac{dx}{L(x)} \\ &= [\ln 2(x-t+L(x))]_{s=0}^{s} = \ln \frac{s-t+L}{1-t} \end{aligned} \tag{3.108}$$

and

$$\begin{aligned} S_4 &= \sum_{n=0}^{\infty} \frac{s^{n+2}}{n+2} P_n(t) = \int_0^s \sum_{n=0}^{\infty} x^{n+1} P_n(t) dx = \int_0^s \frac{x}{L(x)} ds \\ &= [\mathrm{L(x)} + t\ \ln 2(x-t+L(x))]_{x=0}^{s} = L - 1 + t\ \ln \frac{s-t+L}{1-t} \end{aligned} \tag{3.109}$$

Appendix 2: Solutions to Exercises

Solution to Exercise 3.0:

$$\Delta W = \Delta V + \Delta\overline{\Omega} = 2\omega + \left\{ \begin{array}{c} 0 \\ -4\pi\rho \end{array} \right\} \left\{ \begin{array}{c} \text{exterior of the Earth} \\ \text{inside the Earth} \end{array} \right.$$

Solution to Exercise 3.1: Let us introduce the notations g, a and g_Ω for the magnitudes of gravity, gravitation and the centrifugal force. Then the maximum and minimum gravity are given at the poles and the Equator, respectively, with

$$g_{\max} = a - g_\Omega(\theta = 0) = a, \quad \text{where } a \text{ is constantly 981 Gal.}$$

and

$$g_{\min} = a - g_\Omega(\theta = \pi/2)$$

As $g_\Omega(\theta) = \left|\frac{\partial\overline{\Omega}}{\partial r}\right| = R\omega^2 \sin^2\theta$, it follows with R = 6371 km and ω according to Table 1.1, that:

$$g_\Omega(\theta = \pi/2) = 6.371 \times 10^5 \times 7.292^2 \times 10^{-10} = 0.339[\text{Gal}].$$

Hence, gravity at the Equator is 339 mGal less than at the pole for a spherical, homogeneous Earth model. (For this model the difference is only due to the rotation of the model.)

Solution to Exercise 3.2: The gravity, gravitation and centrifugal force vectors (with notations adopting those in the previous exercise) are illustrated in Fig. 3.4a. If the Earth stops rotating, vector $\bar{g}$ will move to and be equal to vector $\bar{a}$. From Fig. 3.4b, the sine theorem can be applied, yielding the equation

$$\frac{\sin\alpha}{g_\Omega} = \frac{\sin\delta}{a} = \frac{\sin(\pi - \varphi - \alpha)}{a} = \frac{\sin(\varphi + \alpha)}{a} = \frac{\sin\varphi\,\cos\alpha + \cos\varphi\,\sin\alpha}{a},$$

which can be simplified to:

$$\tan\alpha = \frac{g_\Omega \sin\varphi}{a - g_\Omega \cos\varphi}.$$

As $\varphi = \pi/3$ and α is a small angel, the equation can be rewritten as:

$$\alpha \approx \tan\alpha = \frac{\sqrt{3}g_\Omega}{2a - g_\Omega}.$$

In this case, $g_\Omega = R\omega^2/4 = 0.085$ [Gal], so that the numerical solution for the change of the plumb line is about ($\alpha \approx$) 15″ towards the north.

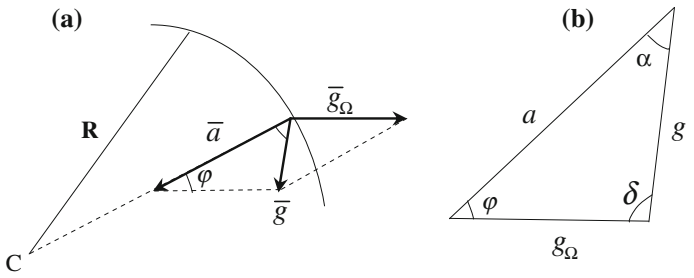

Fig. 3.4 The gravitation, centrifugal force and gravity **a** vectors and **b** their magnitudes in a triangle

Solution to Exercise 3.3 Let us start from the expression

$$r\Delta g = -H - 2T$$

where

$$H = r\frac{\partial T}{\partial r} = \bar{r} \cdot grad(T) = x\frac{\partial T}{\partial x} + y\frac{\partial T}{\partial y} + z\frac{\partial T}{\partial z} = ()T.$$

with

$$() = \left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} + z\frac{\partial}{\partial z}\right).$$

Then

$$\frac{\partial H}{\partial u} = \frac{\partial T}{\partial u} + ()\frac{\partial T}{\partial u}$$

and

$$\frac{\partial^2 H}{\partial u^2} = 2\frac{\partial^2 T}{\partial u^2} + ()\frac{\partial^2 T}{\partial u^2},$$

where $u = x, y, z$.

Hence

$$\Delta H = 2\Delta T + ()\Delta T,$$

and therefore:

$$\Delta(r\Delta g) = -\Delta H - 2\Delta T = 0.$$

Solution to Exercise 3.4 Insert (3.24b) into (3.23b) and apply it for $r_p = R$:

$$\Rightarrow \Delta g = \sum_{n=2}^{\infty} \frac{n-1}{R} T_n.$$

Insert "the hint":

$$\Delta g = \sum_{n=2}^{\infty} \Delta g_n = \sum_{n=2}^{\infty} \frac{n-1}{R} \Omega_n \Delta g_n \Rightarrow \Omega_n = R/(n-1).$$

Hence, by applying "the hint" and (3.23b) once more and changing the order of summation and integration one obtains:

$$T = R \sum_{n=2}^{\infty} \frac{\Delta g_n}{n-1} = \frac{R}{4\pi} \iint_{\sigma} S(\psi) \Delta g d\sigma, \quad \textit{where } S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi).$$

References

Bois GP (1961) Tables of indefinite integrals. Dover Publications Inc., New York

Bruns H (1878) Die Figure der Erde. Publ. Preuss. Geod. Inst, Berlin

Ekman M (1989) Impacts of geodynamic phenomena on systems for height and gravity. Bull Geod 63:281–296

Gauss FW (1828) Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona durch Beobachtungen am Ramsdenschen Zenithsector. Vanderschoeck und Ruprecht, Göttingen, pp 48–50

Heiskanen WA, Moritz H (1967) Physical geodesy. W.H. Freeman and Co., San Francisco

Heiskanen WA, Vening Meinesz FA (1958) The earth and its gravity field. McGraw-Hill Inc

Hotine M (1969) Mathematical geodesy. U. S. Department of Commerce, Washington, D.C

Hwang C (1998) Inverse Vening Meinesz formula and deflection-geoid formula: applications to prediction of gravity and geoid over South China Sea. J Geod 72:304–312

Kaula WM (1966) Theory of satellite geodesy. Blaisdell Publications, London

Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture Notes in Earth Sciences 73. Springer, Berlin, Heidelberg, New York

Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitational field and figure of the Earth. Trans from Russian by Israel Programme for Scientific Translations, Jerusalem

Moritz H (2000) Geodetic reference system 1980. J Geod 74:128–133

Pizzetti P (1911) Sopra il calcolo teorico delle deviazioni del geoide dall' ellipsoide. Atti Accad Sci Torino 46:331

Sjöberg LE (2000) Topographic effects by the Stokes-Helmert method of geoid and quasigeoid determinations. J Geod 74(2):255–268

Sjöberg LE (2014) On the topographic effects by Stokes formula. J Geod Sci 4:130–135

Somagliana C (1929) Teoria generale del campo gravitazionale dell' ellisoide di rotazione. Mem Soc Astron Ital, vol IV

Stokes GG (1849) On the variation of gravity on the surface of the earth. Trans Cambridge Phil Soc 8:672–695

Vening Meinesz FA (1928) A formula expressing the deflection of the plumb-line in the gravity anomalies and some formulae for the gravity field and the gravity potential outside the geoid. Proc Koninkl Ned Akad Wetenschap 31(3):315–331

Chapter 4
Modern Physical Geodesy

Abstract Modern physical geodesy and geophysics employ the surface gravity anomaly or gravity disturbance rather than the classical gravity anomaly defined at sea level. In this way topographic corrections are treated more rigorously. Here the geoid is determined by an Earth gravitational model (EGM) as spherical harmonics, as well as by the modified Stokes' and Hotine's formulas, the latter methods being combinations of surface integrals and an EGM expressed in a series of spherical harmonics. Various types of modifying Stokes' formula are derived, such as methods that reduce the truncation error (alone), remove-restore techniques with or without a higher order reference field, as well as the least squares modification technique. The last method minimizes the errors of truncating the integration area to a spherical cap, as well as the errors stemming from gravity anomalies and the EGM by spectral weighting of the observables. As a result, each of these error contributors as part of the total geoid-error estimate can easily be visualized.

Keywords Modified Hotine's formula · Modified stokes' formula · No-topography gravity anomaly · Topographic bias · Quasigeoid

4.1 Introduction

In 1957 the first artificial satellite was launched to orbit the Earth, enabling for the first time accurate determination of the low-degree harmonics of the Earth's gravity field from dynamic satellite-geodesy observations. This technique started with optical observations, but was gradually improved by refined observation techniques, such as satellite Doppler, laser ranging and GPS, which enabled the determination of higher and higher degree harmonics to increased levels of accuracy. Of particular importance, also was the advent of satellite-radar altimetry, which, for the first time, made possible a detailed mapping of the gravity field over the world's oceans.

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_4

The advance of satellite-derived gravity information can be illustrated as follows (in the spirit of Kaula 1963):

(1) The dynamic satellite orbit analyses techniques from ground-based observations enables the determination of variations of the radius of the satellite level (at mean radius r) from the geopotential difference $\Delta\tilde{W}$ along the satellite orbit according to Bruns' formula:

$$\Delta h = \frac{\Delta\tilde{W}}{\gamma} = r\sum_{n=0}^{n_{\max}}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n} C_{nm}Y_{nm}(\Omega), \tag{4.1}$$

where R is the mean Earth radius. Considering Kaula's rule (Eq. 3.29), this observation has the following power spectrum (degree variance) at satellite level:

$$c_n^2(\Delta h) \approx r^2\left(\frac{R}{r}\right)^{2n+2}\frac{2n+1}{n^4}10^{-10}. \tag{4.2}$$

Only those harmonics resulting in signal-to-noise ratios bigger than one are useful. Assuming that the observation accuracy is 1 m (0.1 m), in the case of ground-based methods, Fig. 4.1 shows that $n_{\max}$ is not larger than, say, 10 (20). However, of

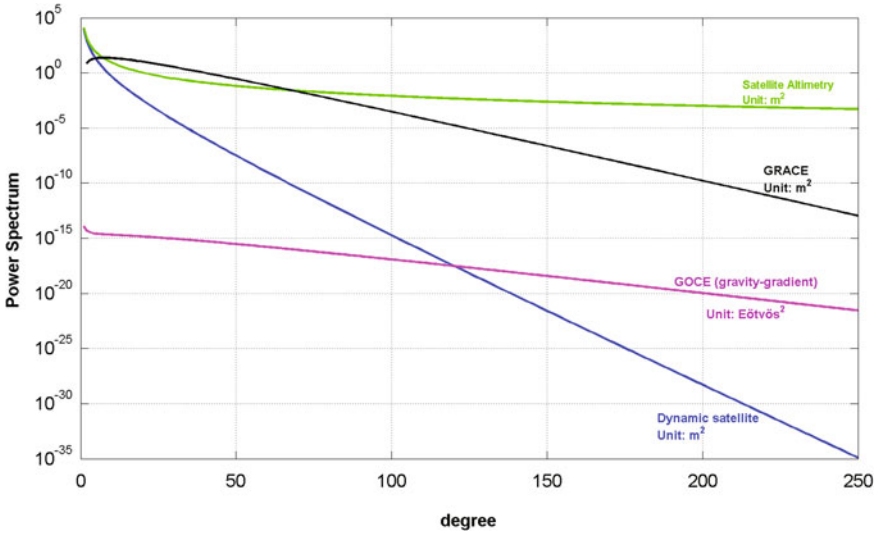

Fig. 4.1 Degree variances c_n^2 using Eqs. (4.2) with r = R + 1000 km, Eqs. (4.3), (4.8) with Δs = 250 km and (4.10) with r = R + 250 km. R = 6371 km. Note that the units are m^2 except for Eq. (4.10), where the unit is $Eötvös^2$

particular importance are long-term satellite laser-ranging data, which are useful for accurate determination of low-degree harmonics and their temporal variations.

(2) The radar in satellite altimetry operates at satellite level, but the technique directly measures (when disregarding some systematic errors; see below) the geoid height at sea, with the power spectrum of:

$$c_n^2(N) = c_n^2(T/\gamma_0) \approx R^2 \frac{2n+1}{n^4} 10^{-10}, \tag{4.3}$$

which is significantly more sensitive to high-degree harmonics than those in Eq. (4.2).
Importantly, the method is practically limited to ocean areas (where terrestrial gravity data is very sparse), and it actually measures mean sea-surface topography, which deviates from the geoid level surface by about $\pm 1 - 2$ m.

(3) Low-low satellite-to-satellite tracking (example: GRACE mission) mainly measures the difference velocity Δv between the absolute velocities v_i squared between the two twin-satellites, and this observation yields the following observation equation based on the energy integral (Wolff 1969; Sjöberg 1982):

$$W_2 - W_1 = \frac{v_2^2 - v_1^2}{2} \approx \bar{v}\Delta v, \tag{4.4}$$

where W_i are the geopotentials at the two satellites, and $\bar{v} \approx \sqrt{GM/r}$ is their mean velocity. As the geopotential is the sum of normal potential and disturbing potential, the potential difference can be approximated by:

$$W_2 - W_1 = U_2 - U_1 + T_2 - T_1 \approx \frac{\partial T}{\partial s}\Delta s = \gamma\bar{\theta}\Delta s, \tag{4.5}$$

where Δs is the separation between the satellites along the orbit arc (s), and $\bar{\theta}$ is the deflection of the vertical along the orbit. (In Eq. (4.5), we assume that $\partial U/\partial \mathrm{s} = 0$, implying that the two satellites are located on a level surface in the normal field for the arc of length Δs.) The satellite-orbit analysis makes possible the following determination of the satellite level variations:

$$\Delta h = \frac{\bar{v}\Delta v}{\gamma} \approx \bar{\theta}\Delta s, \tag{4.6}$$

and the degree variances for the deflection component are given by (Heiskanan and Moritz 1967, Eq. 7.38) as:

$$c_n^2(\bar{\theta}) = \frac{(n-1)^2(n+1)n}{\gamma^2 r^2} c_n^2(T), \tag{4.7}$$

yielding the power spectrum for Δh:

$$
\begin{aligned}
c_n^2(\Delta h) &\approx c_n^2\left(\bar{\theta}_n\right)(\Delta s)^2 \approx \frac{1}{\gamma^2}\frac{(n-1)^2(n+1)n}{r^2}c_n^2(T)\left(\frac{\Delta s}{r}\right)^2 \\
&\approx (\Delta s)^2\left(\frac{R}{r}\right)^{2n+2}\frac{n(n^2-1)(n-1)(2n+1)}{n^4}10^{-10}
\end{aligned}
\tag{4.8}
$$

In case of the GRACE mission, Δs is of the order of 250 km.
Finally, the observation of the vertical component in satellite gradiometry at radius r can be expressed by Eq. (3.68b):

$$
g_r = -\frac{1}{R^2}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+3}(n+1)(n+2)T_n(\Omega), \tag{4.9}
$$

yielding the following gravity-gradient power spectrum:

$$
c_n^2(g_r) \approx \left(\frac{GM}{R^3}\right)^2\frac{(n+1)^2(n+2)^2(2n+1)}{n^4}\left(\frac{R}{r}\right)^{2n+6}10^{-10}. \tag{4.10}
$$

In Fig. 4.1, the above ideal power spectra are compared, showing that the spectrum of the classical dynamic satellite techniques (Eq. 4.1) drops very quickly towards zero for increasing degrees, while the satellite altimetry spectrum, represented by Eq. (4.3), performs exceptionally well, maintaining the power over the range of the spectrum. In case of satellite-to-satellite tracking (Eq. 4.8) and satellite gradiometry (Eq. 4.10), the spectra decrease relatively slowly, revealing there feasibilities to detect much higher harmonics than the classical technique. In fact, as of 2015, the dedicated satellite gravity missions GRACE and GOCE have recovered the low- to medium-wavelength gravity field (say for n < 300) to an unprecedented accuracy.

The uncertainty in recovered geopotential coefficients are also dependent on the accuracy and limitations of the observation system. The actual development is illustrated in Fig. 4.2, which shows the accuracies by wavelength of some well-known EGMs. For comparison, the actual signal degree variances of the geopotential (represented by EGM2008) is also plotted. Since the millennium's beginning, three dedicated satellite-gravity missions (CHAMP, GRACE and GOCE) have very significantly improved our knowledge about the low and medium wavelengths of the gravity field. However, to attain the on-going challenge of "the 1-cm geoid", a combination of satellite-derived gravity data with terrestrial information is needed.

Progress in geoid determination cannot rely solely on data improvement, but theory must also be improved. In 1962, the Russian geodesist M.S. Molodensky and his colleagues (Molodensky et al. 1962) published in the West their remarkable new theory on determining the shape of the Earth from levelling and gravity data without any requirement for knowing the density distribution of the topography,

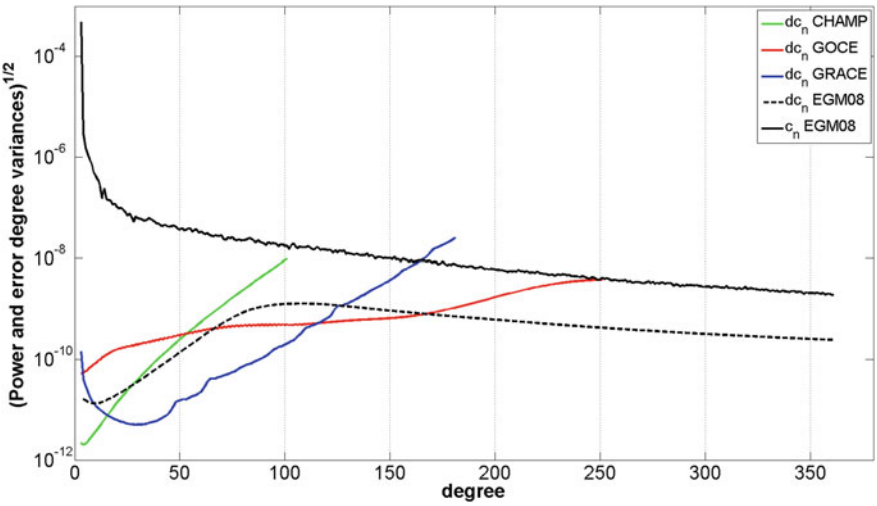

Fig. 4.2 Square root of geopotential power and error-degree variances $(c_n \text{ and } dc_n)$ using CHAMP (*aiub-champ03 s* model up to degree 100; Prange 2011), GOCE (*go_cons_gcf_2_tim_r3* model up to degree 250; Pail et al. 2011), GRACE (*ggm05 s* model up to degree 180; Tapley et al. 2013) and EGM08 (Pavlis et al. 2008) EGM models

which is compulsory in the traditional techniques for determining the geoid and orthometric height. In the new theory, the concepts of orthometric height and geoid height were replaced by the *normal height* and the *quasigeoid height* or height anomaly. In addition, the classical gravity anomaly, located on the geoid, was replaced by *the surface gravity anomaly*.

The above concepts are most important for accurate geoid and quasigeoid determinations. Since the 1960s, most research in physical geodesy follow Molodensky's pioneer work with a focus on quasigeoid estimation (e.g. Bjerhammar 1962, 1963 and Moritz 1980). Even today, textbooks, such as Sansó and Sideris (2013), are primarily concerned with the quasigeoid, despite the fact that many countries around the world (e.g. on the North and South American, African and Australian continents) adopt the geoid for their national reference models.

Already Molodensky et al. (1962, Chap. 7) presented a method to determine the height anomaly by combining satellite-derived, long-wavelength information with detailed gravity data around the computation point, and a variety of such techniques are successfully used today.

In this chapter, we will first deal with modern geoid determination by taking advantage of some of Molodensky's ideas, but only later (in Sect. 7.1) will we present methods for quasigeoid determination. Hence, this chapter is organized as follows: Sect. 4.2 presents the surface and Bouguer gravity anomalies and disturbances. We also define the no-topography disturbing potential and gravity anomaly. Sections 4.3 and 4.4 are devoted to precise geoid determination from spherical harmonics alone

and in combination with Stokes' formula (Stokes 1849). Section 4.5 summarizes the modified Stokes' formulas, while Sect. 4.6 uses the modified Hotine formula with gravity disturbances and an EGM for geoid determination.

4.2 The Quasigeoid, Surface Gravity Anomaly and Disturbance

Definition 4.1 *(the telluroid and quasigeoid):* Along each normal to the reference ellipsoid, there is a point Q whose normal potential U_Q equals the Earth's potential W_P at the topographic surface (see Fig. 4.3). The surface that connects all points Q is called the telluroid. The distance along the normal between points Q and P is called the *height anomaly* or *quasigeoid height* (ζ), and the distance from the reference ellipsoid to Q is called *the normal height* (H^N).The *quasigeoid* is the mapping of the ζ on the reference ellipsoid (see Fig. 4.3).

As can be seen in the figure, the geodetic height h is both the (approximate) sum of the geoid and orthometric heights, as well as the sum of the height anomaly and normal height. The difference between the geoid and quasigeoid heights ranges up to a few metres in high mountains and vanishes on the ocean.

Definition 4.2 *(gravity anomaly and disturbance)*: The surface gravity anomaly and gravity disturbance are given by $\Delta g_P = g_P - \gamma_Q$ and $\delta g_P = g_P - \gamma_P$, respectively.

We notice that the anomaly and disturbance look the same as those in the classical definitions, but here point P and Q are located at the Earth's surface and telluroid, while in the classical definition P is on the geoid and Q is at the reference ellipsoid. In fact, P can be located anywhere in space, and Q is the corresponding point along the ellipsoidal normal through P that obeys $U_Q = W_P$.

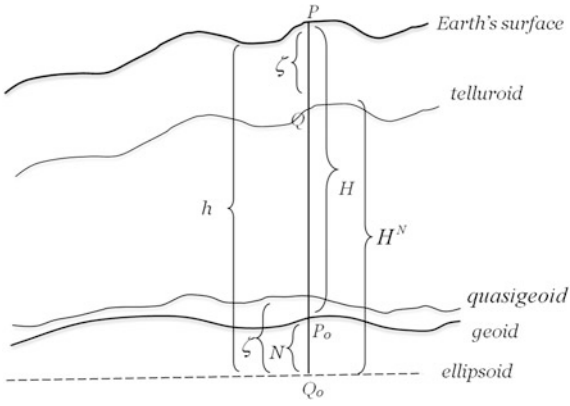

Fig. 4.3 Illustration of the concepts quasigeoid and telluroid

Proposition 4.1 *The quasigeoid height (or height anomaly)* is given by Bruns' formula:

$$\zeta_P = \frac{T_P}{\gamma_Q}. \tag{4.11}$$

Proof $T_P = W_P - U_P = W_P - \left(U_Q + U_Q'\zeta_P + \ldots\right) = (W_P = U_Q) = \gamma_Q\zeta_P + \ldots$ and the proposition follows as a first-order Taylor approximation.

Note that Eq. (4.11) is an approximation, but its relative error is only of the order of $|\zeta/r| \leq 0.02\%$.

Proposition 4.2 *The fundamental equation of physical geodesy*

$$\Delta g_P = -\left(\frac{\partial T}{\partial h}\right)_P + \left(\frac{\partial \gamma}{\partial h}\right)_Q \frac{T_P}{\gamma_Q} \tag{4.12a}$$

and its spherical approximation

$$\Delta g_P = -\left(\frac{\partial T}{\partial r}\right)_P - 2\frac{T_P}{r_P} \tag{4.12b}$$

hold anywhere in space.

Proof By Taylor expansion to the first order, one obtains:

$$\begin{aligned}\left(\frac{\partial T}{\partial h}\right)_P &= -g_P + \gamma_P = -g_P + \left[\gamma_Q + \left(\frac{\partial \gamma}{\partial h}\right)_Q \zeta_P + \ldots\right] \\ &= -\Delta g_P + \left(\frac{\partial \gamma}{\partial h}\right)_Q \frac{T_P}{\gamma_Q} + \ldots,\end{aligned} \tag{4.13}$$

where the magnitude of the next term in the Taylor series, $-3\gamma_Q\zeta_P^2/r_Q^2$, does not exceed 1 μ Gal. Also, $(\partial\gamma/\partial h)_Q$ in Eq. (4.12a) can be substituted by $(\partial\gamma/\partial h)_P$ to a relative error within 0.003%.

One may also define the Bouguer gravity disturbance and anomaly in space as:

$$\delta g^B = \delta g - A^T \text{ and } \Delta g^B = \Delta g - A^T, \tag{4.14}$$

where A^T is the topographic attraction (as in Chap. 3). In the Bouguer gravity disturbance all the topographic signal is removed. However, surprisingly, this is not the case with the Bouguer gravity anomaly. To see this, we decompose the gravity anomaly into the no-topography (NT) and topographic anomalies (Δg^{NT} and Δg^T) by considering the fundamental Eq. (4.12a):

$$\Delta g = \Delta g^{NT} + \Delta g^{T} \;\Rightarrow\; \Delta g^{NT} = \Delta g - \Delta g^{T}, \tag{4.15a}$$

where:

$$\Delta g^{NT} = -\left(\frac{\partial T^{NT}}{\partial h}\right) + \left(\frac{\partial \gamma}{\partial h}\right)\frac{T^{NT}}{\gamma_Q} \tag{4.15b}$$

$$\Delta g^{T} = A^{T} + \left(\frac{\partial \gamma}{\partial h}\right)\frac{V^{T}}{\gamma_Q} \approx A^{T} - 2\frac{V^{T}}{r_Q}. \tag{4.15c}$$

Here the last term of Eq. (4.15c) is the spherical approximation, point Q is the same as in Definition 4.2, V^T is the topographic potential and

$$T^{NT} = T - V^{T} \tag{4.16}$$

is *the no-topography disturbing potential.* We notice that the NT-anomaly needs not only a correction by the direct topographic effect $(-A^T)$ but also by the SITE [i.e. the last term of Eq. (4.15c) with opposite sign]. Hence, the SITE is not a correction needed to change the reference surface in Stokes' integration as assumed in classical physical geodesy (see Sect. 3.4.4), but it is a necessary correction to remove all topographic signal from the gravity anomaly (Sjöberg 2014). It follows that the Bouguer attraction is the total topographic correction for the gravity disturbance, but it is only part of the total topographic effect for the gravity anomaly, and this is because the gravity disturbance in itself (but not the anomaly) is an attraction.

4.3 Geoid Determination by Spherical Harmonics

As discussed in Sect. 3.2.3, Bruns' formula is fundamental for determining the geoid height:

$$N = \frac{T_g}{\gamma_0} = \frac{W_g - U_g}{\gamma_0}. \tag{4.17}$$

In the practical application of Eq. (4.17), the radius of the geoid is usually approximated by the Mean Earth sphere of radius R = 6371 km. However, if the correct radius at the computational point is $r = R + dr$, this approximation implies an error of:

$$dN = -\frac{\delta g}{\gamma_0} dr, \tag{4.18}$$

which could be significant. Hence, one should use the best available radius in the computational area.

As both W_g and U_g include the known rotation potential of the Earth, the disturbing potential is harmonic outside the topography (upon neglecting the atmosphere). Therefore, the geoid height can be estimated by a series of spherical harmonics (complete to degree M):

$$\tilde{N} = \frac{1}{\gamma_0} \sum_{n=0}^{M} \left(\frac{R_0}{r_e} \right)^{n+1} T_n(\theta, \lambda), \tag{4.19a}$$

where:

$$T_n(\theta, \lambda) = \sum_{m=-n}^{n} T_{nm} Y_{nm}(\theta, \lambda), \tag{4.19b}$$

T_{nm} being the potential harmonic coefficients determined as the difference between those of the gravitational potential (represented by an EGM and the normal potential. Here R_0 is the reference radius for the harmonic series (frequently the same as the semi-major axis of the reference ellipsoid) and r_e is the radius of the reference ellipsoid at the computation point. Equation (4.19a) holds only when the geoid is located outside the topography, but, inside the topography, it suffers from the analytical continuation error (see Sect. 5.2.5 for details).

The series in Eq. (4.19a) is not likely to converge within the Brillouin sphere, the bounding sphere including all Earth's topography, at least not inside the topography. However, as expressed by Moritz (2003), the series is probably only asymptotically divergent, implying that it is practical to very high degrees after which it starts to diverge. Some general properties of spherical harmonic series were presented in Sect. 2.5.

The dominant part in the error of analytical continuation (at least for low- to medium-degree EGMs) is the *topographic bias*, which can be expressed for a constant topographic density ρ [see Eq. (5.21)]:

$$dN^T = -\frac{2\pi G\rho}{\gamma_0} \left(H_M^2 + \frac{2}{3} \frac{H_M^3}{R} \right), \tag{4.20}$$

where H_M^i are the Laplace series of H^2 and H^3 to degree M. Possibly it is even more advantageous to use H^2 and H^3 in this formula, which may reach several metres in mountainous regions. Here it is assumed that the density of the topography (ρ) is constant, and the height of the topography is H. Nevertheless, for higher-degree EGMs, the basically diverging series will more and more add an additional significant bias term to Eq. (4.20) (see Sect. 5.2.6).

4.4 The Modified Stokes' Formula

Today, gravity data cover only part of the Earth's surface, while, in principle, Stokes' formula requires complete coverage. In 1960, M.S. Molodensky (see Molodensky et al. 1962) proposed to truncate Stokes' integral to a cap around the computation point and to compensate for the truncation bias by adding a low-degree spherical harmonic series from an EGM. To further reduce the truncation error, Stokes' function is modified in a suitable way. The resulting method is generally named *the* modified Stokes formula. Since Molodensky's pioneer work, various types of modifications to Stokes' formula have evolved, and some of them are described below.

Let us first return to Stokes' original formula in Sect. 3.3.2:

$$N = \frac{c}{2\pi} \iint_{\sigma} S(\psi) \Delta g d\sigma, \tag{4.21a}$$

where we have introduced the constant $c = R/(2\gamma_0)$, and Stokes' function can be expressed as the series

$$S(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} P_n(\cos\psi). \tag{4.21b}$$

Note that Stokes' function starts the summation at $n = 2$, implying that Stokes' integral is blind to any zero- and first-degree harmonics of the gravity anomaly.

4.4.1 General Modification of Stokes' Formula

Taking advantage of the orthogonality of spherical harmonics when integrated over the sphere, it is easy to show that Eq. (4.21a) can also be written as:

$$N = \frac{c}{2\pi} \iint_{\sigma} S^L(\psi) \Delta g d\sigma + c \sum_{n=2}^{L} s_n \Delta g_n, \tag{4.22a}$$

where Δg_n are the so-called Laplace harmonics of Δg defined by (Heiskanen and Moritz 1967, p. 97):

$$\Delta g_n = \frac{2n+1}{4\pi} \iint_{\sigma} \Delta g P_n(\cos\psi) d\sigma, \tag{4.22b}$$

and

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} \frac{2n+1}{2} s_n P_n(\cos\psi) \tag{4.22c}$$

is *the modified Stokes' function*, where L is the selected maximum degree of the arbitrary *modification parameters* s_n. Hence, the modified Stokes' formula yields exactly the same result as the original Stokes' formula, Eq. (4.21a), and nothing is therefore gained so far by the modification.

Let us now introduce a spherical cap σ_0 limited by the spherical angle ψ_0 and centred at the computation point for Stokes' integration. Then it can be shown that Stokes' integration in the exterior area to the cap, $\sigma - \sigma_0$, i.e. the truncation error when omitting the contribution from outside the cap, can be replaced by a Laplace series in Δg, i.e.

$$\frac{c}{2\pi} \iint_{\sigma-\sigma_0} S^L(\psi) \Delta g d\sigma = c \sum_{n=2}^{\infty} Q_n^L(\psi_0) \Delta g_n, \tag{4.23a}$$

where the coefficients are given by:

$$Q_n^L = Q_n^L(\psi_0) = \int_{\psi_0}^{\pi} S^L(\psi) P_n(\cos\psi) \sin\psi d\psi = Q_n(\psi_0) - \sum_{k=2}^{L} E_{nk} s_k \tag{4.23b}$$

$$Q_n(\psi_0) = \int_{\psi_0}^{\pi} S(\psi) P_n(\cos\psi) \sin\psi d\psi \tag{4.23c}$$

and

$$E_{nk} = E_{nk}(\psi_0) = \frac{2k+1}{2} \int_{-1}^{\cos\psi_0} P_n(t) P_k(t) dt. \tag{4.23d}$$

This implies that Eq. (4.22a) can also be written as:

$$N = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi) \Delta g d\sigma + c \sum_{n=2}^{\infty} \left(Q_n^L + s_n^*\right) \Delta g_n, \tag{4.24a}$$

where:

$$s_n^* = \begin{cases} s_n & \text{if} \quad 2 \le n \le L \\ 0 & \text{otherwise} \end{cases}. \tag{4.24b}$$

$Q_n(\psi_0)$, the so-called Molodensky's truncation coefficients, as well as E_{nk}, can be computed by the recurrence relations of Paul (1973) and Hagiwara (1976).

Equation (4.24a) combines gravity anomaly data in the near-zone around the computation point with a spherical harmonic expansion of the gravity field Δg_n by an EGM. For error-free data, this formula yields exactly the same result as the original Stokes' formula, independent of the choice of modification parameters s_n. Importantly, for erroneous data, the modification tunes the weighting between the terrestrial gravity data and the EGM. In practice, however, Eq. (4.24a) must be modified further. First of all, the EGM is known only to a finite upper-degree M, and, as the integration is limited to a spherical cap, this leads to a truncation bias. Second, the true gravity anomaly and spherical harmonic expansion of the gravity field are replaced by estimates Δg^T and Δg^{EGM}, which include various types of errors. Hence, the resulting *general geoid estimator*

$$N_1^{L,M} = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi) \Delta g^T d\sigma + c \sum_{n=2}^{M} \left(Q_n^L + s_n^*\right) \Delta g_n^{EGM} \tag{4.25}$$

will be subject to (a) a truncation error (bias) due to the limited cap-size and/or the limited degree of expansion of the EGM (see Eqs. 4.24a and 4.24b) observation errors ε^T and ε_n^{EGM}, stemming from the errors of the input data of gravity anomaly Δg^T and EGM derived Laplace gravity anomaly Δg_n^{EGM}, respectively. The degree of modification L of Stokes' function is arbitrary. Thanks to the last term of Eq. (4.25), the geoid estimator $N_1^{L,M}$ is unbiased for all degrees 2-M, while the inevitable bias for all higher degrees stems from the truncation. (An estimator of a quantity q is biased if its expected value differs from q. That means that the unbiased estimator approaches its true value as observation errors decrease.) Usually, unbiased estimators are preferred to biased ones. As will be discussed below, by varying the modification parameters s_n, the weights of the input data (and their resulting errors) will vary, leading to different geoid estimators.

Before discussing various special cases of modifications, we will present an even more general geoid estimator, usually biased for all degrees. (It is more general in the sense that the previous estimator $N_1^{L,M}$ is a special case.) The new estimator is given by the formula

$$N_2^{L,M} = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi) \Delta g^T d\sigma + c \sum_{n=2}^{M} b_n \Delta g^{EGM}, \tag{4.26a}$$

or, written in the spectral form, as:

$$N_2^{L,M} = c\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)\Delta g_n^T + c\sum_{n=2}^{M} b_n \Delta g_n^{EGM}. \tag{4.26b}$$

Note that the arbitrary parameters b_n provide further degrees of freedom to this estimator compared to the previous one. It is therefore likely that this model includes an estimator with a smaller error than is available in the geoid model $N_1^{L,M}$. In practice, Eqs. (4.26a, 4.26b) include all published geoid estimators related with the modified Stokes' formula (Sjöberg 2003a, b) (see also Table 4.1). Hence, we are now ready to discuss some of the more common modifications appearing in the geodetic literature.

4.4.2 Remove-Restore Techniques

An early application of using a higher-order reference field in the original Stokes' formula can be found in Vincent and Marsh (1974) and Rapp and Rummel (1975). Their geoid model is a special case (with $L = 0$) of the general geoid estimator for the so-called *remove-compute-restore (RCR) technique:*

$$\tilde{N}_1^{L,M} = \frac{c}{2\pi}\iint_{\sigma_0} S^L(\psi)\Delta g^M d\sigma + c\sum_{n=2}^{M}\frac{2}{n-1}\Delta g_n^{EGM}, \tag{4.27a}$$

where:

$$\Delta g^M = \Delta g^T - \sum_{n=2}^{M}\Delta g_n^{EGM} \tag{4.27b}$$

is the residual gravity anomaly. The rationale behind these estimators is to filter out the low-degree spectrum from Stokes' integral by the high-frequency residual-gravity anomaly data Δg^M, and to represent this part of the spectrum by the long-wavelength harmonics Δg_n^{EGM}. The most typical example is the RCR technique, taught by the International Geoid School (e.g. Forsberg 1993; Sansó 1997; Sansó and Sideris 2013 (see Sect. 8.4.5), where L is sometimes set to zero. Another example is the Vanicek–Kleusberg's geoid estimator, Eq. (4.37a), where $L > M$.

As there is only a high-frequency residual-gravity anomaly under this form of Stokes' integral, it is expected that the truncation error for the far-zone is much smaller than by the corresponding geoid estimator Eq. (4.25). However, as shown by Sjöberg (2005a), this way of arguing is not correct, as the truncation of Stokes'

integral causes leakage of low-frequency geoid harmonics into Stokes' integral. This can be seen by writing Eq. (4.27a) in the partially spectral form:

$$\tilde{N}_1^{L,M} = \frac{c}{2\pi} \iint_{\sigma_0} S^L(\psi) \Delta g^T d\sigma - c \sum_{n=2}^{M} \left[\frac{2}{n-1} - Q_n^L - s_n^* \right] \Delta g_n^{EGM} + c \sum_{n=2}^{M} \frac{2}{n-1} \Delta g_n^{EGM} \tag{4.28}$$

where the first term includes contributions to terrestrial gravity anomaly at all wavelengths from degree 2 and up. Moreover, it follows from Eqs. (4.24a) and (4.28) that $N_1^{L,M}$ and $\tilde{N}_1^{L,M}$ are equal, at least from a theoretical point of view. Hence, the latter geoid estimator is as sensitive to truncation error as the former, and therefore there is no such advantage in using the more complicated geoid estimators with a higher-order reference field. On the other hand, looking from the practical point of view, one may argue that, in the latter technique, the residual gravity anomalies are better suited for interpolating the original data into a grid, to be used in the numerical Stokes' integration. (However, this interpolation can certainly be performed separately without steering the modification of Stokes' formula.) Nevertheless, it should be admitted that the RCR technique, operating with a residual gravity anomaly, could be advantageous from a numerical point of view also in the computation of the downward continuation effect of the gravity anomaly and in the Stokes' integration.

The RCR technique is frequently used also with a removal and restoration of a topographic model, which may reduce also the high-frequency contents of the residual-gravity anomaly, thus making this anomaly even more attractive for interpolation, integration, downward continuation and, in this case, also for the truncation of Stokes' integral. However, as pointed out by Sjöberg (2005a), this possible advantage is easily spoilt if the RCR technique is not performed in a consistent way for the removal and restoration of the effects of the topography.

If $L > M$ and there are potential coefficients available to the higher degree L, one can augment $\tilde{N}_1^{M,M}$ to:

$$N_{VK}^{L,M} = \tilde{N}_1^{M,M} + c \sum_{n=M+1}^{L} Q_n^M \Delta g_n^{EGM} \tag{4.29}$$

This estimator will be further discussed in the next section [see Eqs. (4.37a)–(4.37c)].

4.4.3 Modifications Reducing the Truncation Error

When limiting the area of integration of the original Stokes' formula to a cap σ_0 with geocentric angle ψ_0, the truncation error becomes:

$$\delta N_1^{0,0} = N_1^{0,0} - N = -\frac{c}{2\pi} \iint\limits_{\sigma-\sigma_0} S(\psi)\Delta g d\sigma = -c\sum_{n=2}^{\infty} Q_n \Delta g_n \tag{4.30}$$

which leads to a very significant truncation bias unless the integration area is very large. For instance, for a cap size of 20°, the bias is of the order of 17 m. By selecting the simply modified Stokes' formula

$$N_1^{0,M} = \frac{c}{2\pi} \iint\limits_{\sigma_0} S(\psi)\Delta g d\sigma + c\sum_{n=2}^{M} Q_n \Delta g_n^{EGM} \tag{4.31}$$

with M sufficiently high, the truncation error

$$\delta N_1^{0,M} = -c\sum_{n=M+1}^{\infty} Q_n \Delta g_n \tag{4.32}$$

is much reduced compared to the result of Eq. (4.32). However, for M = 360, a cap size of 20° is still needed to reduce the truncation error to below 1 cm. As argued above, $N_1^{0,M}$ is used in practice in the original RCR technique. (In the RTM method of Forsberg (1984), the truncation error is somewhat reduced by the removal and restoration of the gravity signal of a topographic model.) A suitable modification of Stokes' function can reduce this bias considerably.

A well-known method of modifying Stokes' formula is named after Wong and Gore (1969). It uses the general estimator in Eq. (4.24a) with $L = M$ and parameters s_n set to $2/(n-1)$, which has the effect that Stokes' function is truncated at degree M to a so-called *spheroidal* kernel. The rationale behind this modification is similar to that adopted in the RCR technique, namely to filter out the long-wavelength contribution from the terrestrial data and to fully compensate for that spectrum by the EGM. However, as already stated, the filtering out of the long-wavelengths is not perfect for a truncated integration. Nevertheless, the Wong-Gore modification is a more efficient high-pass filter than the spherical Stokes' kernel (Vaníček and Featherstone 1998). In this case, the truncation error can be written as:

$$\delta N_{WG}^{M} = -c\sum_{n=M+1}^{\infty} Q_n^M \Delta g_n \tag{4.33}$$

which yields a smaller bias than Eq. (4.32) thanks to the fact that the spheroidal Stokes' kernel implies that the long wavelengths are excluded, which is therefore accompanied by smaller-valued truncation coefficients Q_n^M than Q_n, provided by the original Stokes' kernel (see the figures in Vaníček and Featherstone 1998). Another idea of Wong and Gore (1969) was to reduce the truncation bias by choosing the cap size ψ_0 at the point where the residual Stokes' function $S^M(\psi_0)$ is zero. Heck and Grüninger (1987) extended this approach by introducing Meissl's (1971) idea

into the Wong-Gore function, namely to use the [further] modified Stokes' kernel $S^M(\psi) - S^M(\psi_0)$ for any cap size. Alternatively, the degree M of Wong–Gore modification can be chosen such that the zero crossing points of the modified kernel coincide with the desired cap radius. Whatever variant of this approach is taken, it ultimately results in an accelerated rate of convergence of the truncation bias (Evans and Featherstone 2000). Finally, it should also be mentioned that the NKG geoid models from 2004 to 2016, based on a Danish RCR method, used a Wong–Gore type modification of Stokes function. (In 2016, the official NKG geoid model was determined by the Least Squares Modification of Stokes' formula; see Chap. 6).

The method of M.S. Molodensky (Molodensky et al. 1962, pp. 150–152) is probably the original method of modifying Stokes' formula. Moreover, in contrast to all methods presented so far (with an exception for the method of Vanicek and Kleusberg 1987, which is a generalization of Molodensky's method, see below), it is the only method that is based on a criterion to minimize the truncation error. The derivation of the modification starts from the general formula Eq. (4.24a) for $L = M$ written in the form:

$$N = \frac{c}{2\pi} \iint_{\sigma_0} S^M(\psi)\Delta g d\sigma + \frac{c}{2\pi} \iint_{\sigma-\sigma_0} S^M(\psi)\Delta g d\sigma + c\sum_{n=2}^{M} s_n \Delta g_n \tag{4.34}$$

where the intermediate term on the right hand-side is the minus of the truncation bias. By implying Schwartz inequality, it can be shown (Sjöberg 1984a, b) that the upper limit of this bias is a minimum, if $\iint_{\sigma-\sigma_0} (S^M)^2 d\sigma = \min.$ with respect to the modification parameters s_n. This leads to the system of equations

$$\sum_{k=0}^{M} E_{nk} s_k = Q_n \text{ for } n = 0, 1, \ldots, M \tag{4.35a}$$

or, with reference to Eq. (4.23b),

$$Q_n^M = 0 \text{ for } n = 0, 1, \ldots, M, \tag{4.35b}$$

and the resulting geoid estimator becomes a special case of both $N_1^{M,M}$ and $N_2^{M,M}$:

$$N_{Mol.}^M = N_1^{M,M} = N_2^{M,M} = \frac{c}{2\pi} \iint_{\sigma_0} S^M(\psi)\Delta g d\sigma + c\sum_{n=2}^{M} s_n \Delta g^{EGM} \tag{4.36}$$

where the parameters b_n of $N_2^{M,M}$ are equal to s_n, as given in Eq. (4.35a).

Another geoid estimator is the adaptation of Molodensky's technique for a higher-order reference field, as opposed to the above solution (Vaníček and Kleusberg 1987; Vaníček and Sjöberg 1991):

$$N_{VK}^{L,M} = \frac{c}{2\pi}\iint_{\sigma_0} S^M(\psi)\Delta g^M d\sigma + c\sum_{n=2}^{M}\frac{2}{n-1}\Delta g_n^{EGM} + c\sum_{n=M+1}^{L} Q_n^M \Delta g_n^{EGM} \quad (4.37a)$$

where $L \geq M$. The modified kernel is now given by:

$$S^M(\psi) = S(\psi) - \sum_{n=2}^{M}\frac{2n+1}{2}\left(\frac{2}{n-1} + t_n\right)P_n(\cos\psi), \quad (4.37b)$$

and the parameters t_n, which are functions of ψ_0, are determined from the system of equations

$$\sum_{k=2}^{M} E_{nk}t_k = Q_n - \sum_{k=2}^{M}\frac{2}{k-1}E_{nk} = Q_n^M;\ \mathrm{n} = 2,3,\ldots,\mathrm{M}. \quad (4.37c)$$

However, as discussed above, the Vaníček-Kleusberg geoid estimator can also be written in the form of Eq. (4.29), and for $L = M$ the estimator equals Molodensky's estimator. For $L > M$, the modification parameters are those of Molodensky's estimator for $2 \leq n \leq M$, and the estimator is unbiased through degree L.

Further modifications for reducing the truncation bias can be found, e.g. in Vanicek and Featherstone (1998), Evans and Featherstone (2000) and Sjöberg and Hunegnaw (2000).

4.4.4 *The Least Squares Modification of $N_1^{L,M}$ and $N_2^{L,M}$*

It is well known that the errors of the methods discussed in Sect. 4.4.2 are rather strongly dependent on the choice of cap size, and frequently the optimum cap size of a specific method is sought with the goal of yielding least error. The fact that there exists an optimum cap size suggests that the method of modification does not account for all significant sources of errors. In particular, errors stemming from the potential coefficients and the gravity anomalies, so far neglected, should also be considered in the optimization process. The treatment of all these error sources in a least squares sense is the intention of this section. To do so, one has to assume that the error-variance spectra of both the terrestrial data and the EGM are known. The derivations are given next.

First, we assume that the spectral errors ε_n^T and ε_n^{EGM} of the respective observables Δg^T and Δg_n^{EGM} are random with expectation zero, i.e. $\Delta g_n^T = \Delta g_n + \varepsilon_n^T$ and $\Delta g_n^{EGM} = \Delta g_n + \varepsilon_n^{EGM}$, where Δg_n is the true Laplace harmonic of the gravity

anomaly, and we rewrite the general geoid estimator $N_2^{L,M}$ of Eq. (4.26a) in the spectral form

$$N_2^{L,M} = c\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)\left(\Delta g_n + \varepsilon_n^T\right) + c\sum_{n=2}^{M} b_n\left(\Delta g_n + \varepsilon_n^{EGM}\right) \tag{4.38}$$

As the spectral form of the true geoid undulation is:

$$N = c\sum_{n=2}^{\infty}\frac{2\Delta g_n}{n-1}, \tag{4.39}$$

it follows that the error of $N_2^{L,M}$ can be written as:

$$\delta N_2^{L,M} = c\sum_{n=2}^{\infty}(b_n^* - s_n^* - Q_n^L)\Delta g_n + c\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - s_n^* - Q_n^L\varepsilon_n^T\right) + c\sum_{n=2}^{M} b_n\varepsilon_n^{EGM}, \tag{4.40a}$$

where:

$$b_n^* = \begin{cases} b_{n,} & \text{if} \quad 2 \le n \le M \\ 0 & \text{otherwise} \end{cases}. \tag{4.40b}$$

Introducing the global error degree-variances for Δg^T and Δg^{EGM}:

$$\sigma_n^2 = E\left\{\frac{1}{4\pi}\iint_{\sigma_0}(\varepsilon_n^T)^2 d\sigma\right\} \tag{4.41a}$$

and

$$dc_n^2 = E\left\{\frac{1}{4\pi}\iint_{\sigma}(\varepsilon_n^{EGM})^2 d\sigma\right\}, \tag{4.41b}$$

where E{} is the statistical expectation operator, and also introducing the notation c_n^2 for the gravity anomaly degree-variance, i.e.

$$c_n^2 = \frac{1}{4\pi}\iint_{\sigma}\Delta g_n^2 d\sigma, \tag{4.42}$$

and, assuming that all observation errors are random with expectation zero and mutually uncorrelated, one arrives at the following spectral form of *the expected global mean square error* of the geoidal undulation estimator of Eq. (4.26a):

$$\left(\delta\bar{N}_2^{L,M}\right)^2 = E\left\{\frac{1}{4\pi}\iint_{\sigma}\left(\delta N_2^{L,M}\right)^2 d\sigma\right\}$$
$$= c^2\sum_{n=2}^{\infty}\left[\left(b_n^* - s_n^* - Q_n^L\right)^2 c_n^2 + \left(\frac{2}{n-1} - s_n^* - Q_n^L\right)^2\sigma_n^2\right] + c^2\sum_{n=2}^{M} b_n^2 dc_n^2. \tag{4.43}$$

The expected mean square error is thus the sum of the variances and the bias squared, where the sum of variances stems from observation errors and the bias from the truncation of the integration area to a cap and the EGM to degree and order M. By selecting the parameters s_n and b_n in a suitable way, it is possible to minimize the mean square error. Mathematically, by differentiating Eq. (4.43) with respect to each of the parameters b_n and equating to zero, the following least squares solution for b_n (as a function of parameters s_n) is obtained:

$$\hat{b}_n = \left(Q_n^L + s_n^*\right)\frac{c_n^2}{c_n^2 + dc_n^2} \text{ for } 2 \le n \le M. \tag{4.44}$$

Then, by differentiating Eq. (4.43) by each parameter s_n, equating to zero and taking advantage of Eq. (4.44), one arrives at the following system of equations for s_n:

$$c^{-2}\frac{\partial\left(\delta\bar{N}_2^{L,M}\right)^2}{\partial s_k} = 2\sum_{n=2}^{\infty}\left(\delta_{nk} + \frac{\partial Q_n^L}{\partial s_k}\right)\left[\left(Q_n^L + s_n^*\right)C_n - p_n\right] = 0 \tag{4.45a}$$

where:

$$p_k = \frac{2\sigma_k^2}{k-1} \tag{4.45b}$$

$$\delta_{kr} = \begin{cases} 1, & \text{if} \quad k = r \\ 0 & \text{otherwise} \end{cases} \tag{4.45c}$$

and

$$C_k = \sigma_k^2 + \begin{cases} c_k^2 dc_k^2/(c_k^2 + dc_k^2), & \text{if} \quad 2 \le k \le M \\ c_k^2, & \text{if} \quad k > M \end{cases} \tag{4.45d}$$

Considering also that

$$Q_n^L = Q_n - \sum_{k=2}^{L} E_{nk} s_k \tag{4.46a}$$

with

$$\frac{\partial Q_n^L}{\partial s_k} = -E_{nk}, \tag{4.46b}$$

the following system of equations for the parameters s_n follows:

$$\sum_{r=2}^{L} a_{kr} s_r = h_k; k = 2, 3, \ldots, L \tag{4.47a}$$

where:

$$a_{kr} = \sum_{n=2}^{\infty} E_{nk} E_{nr} C_n + \delta_{kr} C_r - E_{kr} C_k - E_{rk} C_r \tag{4.47b}$$

$$h_k = p_k - Q_k C_k + \sum_{n=2}^{\infty} (Q_n C_n - p_n) E_{nk}. \tag{4.47c}$$

We now show that this choice of parameters s_n and b_n correspond to a minimum mean square error solution. For example, by considering Eq. (4.43), the mean square error can be written in the matrix form

$$\left(\delta \bar{N}_2^{L,M}\right)^2 = f + c^2 \mathbf{s}^{\mathbf{T}} \mathbf{A} \mathbf{s} - 2c^2 \mathbf{s}^T \mathbf{h} \tag{4.48a}$$

with f being the mean square error for s_n set to zero:

$$f = c^2 \left[\sum_{n=2}^{\infty} \left(\frac{2}{n-1} - Q_n \right)^2 \sigma_n^2 + \sum_{n=2}^{M} Q_n^2 \frac{c_n^2 d c_n^2}{c_n^2 + d c_n^2} + \sum_{n=M+1}^{\infty} Q_n^2 c_n^2 \right] \tag{4.48b}$$

and matrix $\mathbf{A}$ and vector $\mathbf{h}$ having the elements a_{ij} and h_i, respectively. Then, substituting vector $\mathbf{s}$ by $\hat{\mathbf{s}} + \Delta\mathbf{s}$, where, according to Eq. (4.47a)

$$\hat{\mathbf{s}} = \mathbf{A}^{-1} \mathbf{h} \tag{4.49}$$

and $\Delta\mathbf{s}$ is an arbitrary (but compatible with s) vector, it follows that Eq. (4.48a) can be written as:

$$\left(\delta \bar{N}_2^{L,M}\right)^2 = f - c^2 \hat{\mathbf{s}}^{\mathbf{T}} \mathbf{h} + c^2 \Delta \mathbf{s}^{\mathbf{T}} \mathbf{A} \Delta \mathbf{s} \tag{4.50a}$$

and, as $\mathbf{A}$ is positive definite, this equation proves that the mean square error attains it minimum for $\Delta\mathbf{s} = \mathbf{0}$, and the minimum mean square error becomes:

$$\left(\delta\hat{N}_2^{L,M}\right)^2 = f - c^2\sum_{k=2}^{L}\hat{s}_k h_k. \tag{4.50b}$$

The least squares solution derived here, $\hat{N}_2^{L,M}$, we denote *the optimum modification of Stokes' formula*. As other modifications are special cases of its generating model $N_2^{L,M}$, their expected global mean square errors cannot be less than that of $\hat{N}_2^{L,M}$.

Similarly, the least squares solution for model $N_1^{L,M}$ of Eq. (3.6) becomes:

$$\hat{N}_1^{L,M} = \frac{c}{2\pi}\iint_{\sigma_0} S^L(\psi)\Delta g^T d\sigma + c\sum_{n=2}^{M}\left(Q_n^L + s_n^*\right)\Delta g_n^{EGM}, \tag{4.51a}$$

where the modification parameters are again given by the system of Eqs. (4.47a), but now with:

$$a_{kr} = \delta_{kr}d_k - E_{kr}d_k - E_{rk}d_r + \sum_{n=2}^{\infty}E_{nk}E_{nr}d_n \tag{4.51b}$$

$$h_k = p_k - Q_k d_k + \sum_{n=2}^{\infty}(Q_n d_n - p_n)E_{nk}, \tag{4.51c}$$

where:

$$d_n = \sigma_n^2 + dq_n^2 \tag{4.51d}$$

and

$$dq_n^2 = \begin{cases} dc_n^2 & \text{if} \quad n \le M \\ c_n^2 & \text{otherwise} \end{cases} \tag{4.51e}$$

The mean square error of this estimator is formally the same as in Eq. (4.43), but now with b_n fixed to $Q_n^L + s_n^*$. It can also be formulated by Eq. (4.50b) with:

$$f = c^2\sum_{n=2}^{\infty}\left[\left(\frac{2}{n-1} - Q_n\right)^2\sigma_n^2 + Q_n^2 dq_n^2\right]. \tag{4.52}$$

Although theoretically inferior, from a numerical point of view, the least squares solution for $N_1^{L,M}$ does not deviate much from that for $N_2^{L,M}$ for small cap sizes. However, both least squares solutions minimize their global mean square errors. In areas with small errors in the EGM, $\hat{N}_1$ may be advantageous to $\hat{N}_2$ and vice versa in areas with large EGM errors. Also, as the global mean square error may vary considerably from a local mean square error, Sjöberg (2005a, b) derived a least

squares solution that accounts for such local variations. See also Sjöberg (2011). As the local solutions consider the full covariance matrix of the EGM, it becomes a huge computational task for a high-degree EGM.

Exercise: Let $\sigma_0 \to \sigma$ in (4.51a) and determine the limiting estimator $\hat{N}_1^{M,M}$ ["spectral combination"; see Sjöberg (1981, 1986) and Wenzel (1981, 1982)].

Note. In spectral combination, the modification parameters (s_n) are given directly without solving a system of equations (see Table 4.1).

4.4.5 Satellite Only Low Degree Modifications

From recent and future dedicated satellite gravity and gradiometry missions, it is expected that homogeneous low- and medium-degree EGMs, e.g. to degree and order 300, will be produced to unprecedented accuracies. One can expect that these models will be much less biased and tilt-prone than previous EGMs. Taking such information into consideration, it might be reasonable to design the modification of Stokes' formula in such a way, that the EGM totally determines the corresponding low-degree geoid wavelengths. Starting from the general geoid model $N_2^{L,M}$ written in its spectral form

$$N_2^{L,M} = c\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)\Delta g_n^T + c\sum_{n=2}^{M} b_n \Delta g_n^{EGM}, \tag{4.53}$$

the low-degree satellite-only modifications result from the first bracket set to zero, yielding the following equations to solve for s_n:

$$s_n^* - \sum_{r=2}^{L} E_{nr} s_r = \frac{2}{n-1} - Q_n; n = 2, 3, \ldots, M \tag{4.54}$$

If $L > M$, the solution is not unique, because there are only $M-1$ equations to determine $L-1$ parameters s_n. The general solution for $L \geq M$ can thus be written:

$$N_3^{L,M} = c\sum_{n=2}^{M} b_n \Delta g_n^{EGM} + c\sum_{n=M+1}^{\infty}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)\Delta g_n^T \tag{4.55}$$

with the expected global mean square error:

$$\begin{aligned}\left(\delta\bar{N}_3^{L,M}\right)^2 = {} & c^2\sum_{n=2}^{M}\left[b_n^2 dc_n^2 + \left(b_n - \frac{2}{n-1}\right)^2 c_n^2\right] \\ & + c^2\sum_{n=M+1}^{\infty}\left[\left(Q_n^L + s_n^*\right)^2 c_n^2 + \left(\frac{2}{n-1} - Q_n^L - s_n^*\right)^2 \sigma_n^2\right]\end{aligned} \tag{4.56}$$

The optimum solution for b_n and s_n, minimizing $(\delta\bar{N}_3^{L,M})^2$, are given by:

$$\hat{b}_n = \frac{2}{n-1}\frac{c_n^2}{c_n^2 + dc_n^2}; n = 2, 3, \ldots, M, \tag{4.57}$$

and s_n is given by the system of equations

$$\sum_{r=2}^{L} a_{kr} s_r = h_k \; k = 2, 3, \ldots, L, \tag{4.58a}$$

where:

$$a_{kr} = \begin{cases} \delta_{kr} - E_{kr} & if \quad 2 \le k \le M \\ \delta_{kr} d_k - E_{kr} d_k + \sum\limits_{n=M+1}^{\infty} E_{nk} E_{nr} d_n & if \quad M < k \le L \end{cases} \tag{4.58b}$$

$$h_k = \begin{cases} \frac{2}{k-1} - Q_k & if \quad 2 \le k \le M \\ p_k - Q_k d_k + \sum\limits_{n=M+1}^{\infty} E_{nk}(Q_n d_n - p_n) & if \quad M < k \le L \end{cases} \tag{4.58c}$$

$$d_n = c_n^2 + \sigma_n^2 \text{ and } p_n = 2\sigma_n^2/(n-1) \tag{4.58d}$$

with the expected mean square error

$$(\delta\hat{N}_3^{L,M})^2 = f - c^2 \sum_{n=M+1}^{L} \hat{s}_n h_n \tag{4.59a}$$

where:

$$f = c^2 \sum_{n=2}^{M} \left(\frac{2}{n-1}\right)^2 \frac{c_n^2 dc_n^2}{c_n^2 + dc_n^2} + c^2 \sum_{n=M+1}^{\infty} \left[\left(\frac{2}{n-1} - Q_n^L\right)^2 \sigma_n^2 + \left(Q_n^L\right)^2 c_n^2\right]. \tag{4.59b}$$

The practical form of the optimum solution can thus be written as:

$$\hat{\hat{N}}_3^{L,M} = \frac{c}{2\pi} \iint\limits_{\sigma_0} S^L(\psi) \Delta g \, d\sigma + c \sum_{n=2}^{M} \frac{2}{n-1} \frac{c_n^2}{c_n^2 + dc_n^2} \Delta g_n^{EGM} \tag{4.60}$$

with the mean square error provided by Eqs. (4.59a and 4.59b). This solution is biased for all degrees. A corresponding unbiased least squares estimator (unbiased through degree M) is given by the slight modification to Eq. (4.60) by substituting

the factors $c_n^2/(c_n^2 + dc_n^2)$ by $\hat{s}_n$. The removal of the bias is then gained at the price of a slightly increased mean square error.

Numerical tests by Ågren and Sjöberg (2004) have revealed that the above types of satellite-only modifications are prone to rather large truncation errors. In other words, the restriction of the low-degree parameters s_n of these models to be contributed to only from the EGM leaves too few degrees of freedom to sufficiently reduce the truncation bias.

4.4.6 Modifications with High-Degree EGMs

Already today, there are some ultra-high EGMs for geoid determination, and the question we raise here is how to combine such a model with regional, detailed gravity data. It is obvious that Stokes' modified formula could be used also here in one way or another. However, so far we have assumed that the EGM is included only to a moderate degree of M, being less or equal to the degree (L) of modification of Stokes' function. Theoretically, there is no problem to extend this assumption to high degrees, but the determination of the modification parameters, and the application of the modified Stokes' formula will be formidable and impractical. Hence a more practical solution would be to keep L at a rather low degree and let M go to a very high degree. It is obvious that the previous general unbiased and biased estimators $\tilde{N}_1^{L,M}$ and $\tilde{N}_2^{L,M}$ hold also in this case, and they can be optimized by the least squares choice of modification parameters as before.

One should also remember that the truncation error in the modified Stokes' formula decreases with the cap size, as well as the maximum degree (M) of the harmonic series. Consequently, if M is very high, the truncation error should be small, and, if it is neglected, the least squares modified Stokes' formula becomes the least squares spectral combination (see Table 4.1), which does not need a special computational module for determining the modification parameters (s_n); see also Sjöberg (1981, 1986).

However, there is another aspect that one should consider when using an ultra-high degree EGM, namely that it is rather likely that some or much of the regional terrestrial data to be used in Stokes' integral is already included in the EGM, and this leads to correlation between the terrestrial and EGM data. This problem was already considered by Sjöberg (1984b).

4.5 Summary of Modified Stokes' Formula Techniques

The general geoid estimator of Eq. (4.26a) has been shown to include most methods applied today for modifying Stokes' formula. This is illustrated in Table 4.1, where the parameters are specified for several special cases. The solution of Eq. (4.26a, 4.26b), with the optimum choice of modification parameters according to Eqs. (4.44) and (4.47a–4.47c), should therefore be competitive to any of these alternative methods, including the least squares solutions by Sjöberg (1984a, b, 1991). These optimizations were primarily based on the assumption of no systematic errors among the data, and it is therefore expected that the data are cleansed of such errors prior to Stokes' integration. However, one can assume that terrestrial gravity data are biased at least in the long-wavelength spectrum, which speaks in favour of adopting either of the EGM-only low-degree modifications developed in Sect. 4.4.5. On the other hand, the latter type of geoid models are prone to large truncation biases from the neglected remote zone, and a better choice might be a least squares estimator with a proper down weighting for all errors of the terrestrial gravity data.

The least squares modifications already presented were all based on minimizing the global mean square error. As an alternative, Sjöberg (2005b) presented a technique for local least squares optimization. Unfortunately, the formulas then

Table 4.1 A comparison of the parameters of some methods of modifying Stokes' formula

Method	L	M	b_n	a_{nr}	s_n	h_n
Molodensky	M	Arbitrary	s_n	E_{nr}		Q_n
Meissl (1971)	0	0			$S(\psi_c)$	
Wong and Gore (1969)	M	Arbitrary	$Q_n^M + s_n$		$2/(n-1)$	
RCR-method	0	Arbitrary	Q_n			
Spectral combination	M	Arbitrary	s_n		$\frac{2}{n-1}\frac{\sigma_n^2}{\sigma_n^2+dc_n^2}$	
Sjöberg (1986)	∞	Arbitrary	$\frac{2}{n-1}u_n$		$\frac{2}{n-1}v_n$	
Vanicek and Kleusberg (1987)	$L \le M$	Arbitrary	s_n $2 \le n \le L$	E_{nr}		Q_n
Vanicek and Kleusberg (1987)			Q_n^L $L < n \le M$			
$\tilde{N}_1(unbiased)$	Arbitrary	Arbitrary	$Q_n^L + s_n^*$		Arbitrary	
$\tilde{N}_2(biased)$	Arbitrary	Arbitrary	Arbitrary		Arbitrary	

The notations refer to the general model of Eq. (4.26a). For the least squares methods, the parameters a and h are presented for determining modification parameters s_n from Eq. (4.47a).

(Sjöberg 2003b.) Further explanations: s_n^* was defined in Eq. (4.24b), $u_n = c_n^2\sigma_n^2/p_n$ and $v_n = \begin{cases} c_n^2 dc_n^2/p_n & if \quad 2 \le n \le L \\ c_n^2/(c_n^2+\sigma_n^2) & if \quad n > L \end{cases}$ with $p_n = c_n^2 dc_n^2 + c_n^2\sigma_n^2 + dc_n^2\sigma_n^2$

become much more cumbersome, but this method can probably be further developed and applied with modern computer technology.

The least squares geoid estimators rely on the gravity anomaly signal and observation error spectra. Frequently, it is argued that these spectra are too poorly known to warrant the application of stochastic geoid estimators. Nevertheless, it is usually better to use a coarse error model, than to assume that the data are without errors, as is implicitly the case in most deterministic models. (Admittedly simple Stokes' kernel modification, e.g. according to Wong and Gore's technique, have some of the properties of the least squares modification, but without being based on an optimization criterion.) Also, the computational burden for the optimum modification is not significantly different from other methods of modifying Stokes' formula, because when the cap size, as well as the signal and error spectra are designed, the modification parameters must only be determined once from a set of equations with a positive definite design matrix.

We conclude this section by summarizing some well-known methods of modifying Stokes' formula in Table 4.1. They are all special cases of the general estimator given by Eq. (4.26a) (see Sjöberg 2003b).

4.6 The Modified Hotine Formula

Today the 3-D positions of gravity measurements are frequently provided by satellite positioning, such as by GPS. As the geodetic height is known, one can therefore determine the gravity disturbance rather than the gravity anomaly, and, consequently, Hotine's formula can replace Stokes' formula for geoid determination. The modified Hotine formula is completely analogous to the modified Stokes' formula with the spectral factor $1/(n-1)$ replaced by $1/(n+1)$ everywhere. For example, the general Hotine formula, corresponding to Eq. (4.26a), reads:

$$N_H^{L,M} = \frac{c}{2\pi} \iint_\sigma H^L(\psi)\delta g d\sigma + c\sum_{n=2}^{M} b_n \delta g_n^{EGM}, \tag{4.61a}$$

where $H^L(\psi)$ is the modified Hotine function given by:

$$H^L(\psi) = H(\psi) - \sum_{n=2}^{L} \frac{2n+1}{2} h_n P_n(\cos\psi). \tag{4.61b}$$

Here h_n are arbitrary modification parameters and

$$H(\psi) = \upsilon^{-1} - \ln\left(1 + \upsilon^{-1}\right), \text{ where } \upsilon = \sin(\psi/2), \tag{4.61c}$$

Featherstone (2013) reviewed various methods of modifying Hotine's formula.

References

Ågren J, Sjöberg LE (2004) Comparisons of some methods for modifying Stokes' formula in the GOCE era.In: Proceedings of the second international GOCE user workshop "GOCE, the geoid and oceanography", ESA-ESRIN, Frascati, Italy, 8–10 Mar 2004 (ESA SP-569, June, 2004)

Bjerhammar A (1962) On an explicit solution of the gravimetric boundary value problem for an ellipsoidal surface of reference. Tech rep, The Royal Institute of Technology, Division of Geodesy, Stockholm

Bjerhammar A (1963) A new theory of gravimetric geodesy division of Geodesy. Royal Institute of Technology, Stockholm

Evans JD, Featherstone WE (2000) Improved convergence rates for the truncation error in gravimetric geoid determination. J Geodesy 74(2):239–248

Featherstone WE (2013) Deterministic, stochastic, hybrid and band-limited modifications of Hotine's integral. J Geodesy 87(5):487–500

Forsberg R (1984) A study of terrain reductions, density anomalies and geophysical inversion methods in gravity field modeling, Dept Geod Sci rep No 355. The Ohio State University, Columbus, Ohio

Forsberg R (1993) Modelling the fine-structure of the geoid: methods, data requirements and some results. Surv Geophys 14:403–418

Hagiwara Y (1976) A new formula for evaluating the truncation error coefficient. Bull Geod 50:131–135

Heck B, Grüninger W (1987) Modification of Stokes's integral formula by combining two classical approaches. IUGG General Assembly, Vancouver

Heiskanen WH, Moritz H (1967) Physical geodesy. WH Freeman, San Francisco

Kaula WM (1963) Determination of the earth's gravitational field. Rev Geophys 1:507–551

Meissl P (1971) Preparation for the numerical evaluation of second order Molodensky-type formulas. Dept Geodetic Science Rep No 163, Ohio State Univ, Columbus, Ohio

Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitational field and figure of the Earth. Trans from Russian by Israel Programme for Scientific Translations, Jerusalem

Moritz H (1980) Advanced physical geodesy. Wichmann, Karlsruhe

Moritz H (2003) The strange behaviour of asymptotic series in mathematics, celestial mechanics and physical geodesy. In: Grafarend EW, Krumm FW, Schwarze VS (Hrsg.) Geodesy: the challenge of the third millennium. Springer, Berlin, S. 371–377. (Im Internet: Google-Suche nach "asymptotic series".)

Pail R, Bruinsma S, Migliaccio F, Foerste C, Goiginger H, Schuh W-D, Hoeck E, Reguzzoni M, Brockmann JM, Abrikosov O, Veicherts M, Fecher T, Mayrhofer R, Krasbutter I, Sanso F, Tscherning CC (2011) First GOCE gravity field models derived by three different approaches. J Geodesy 81:11. doi:10.1007/s00190-011-0467-x

Pavlis N, Holmes SA, Kenyon SC, Factor JK (2008) An earth gravitational model to degree 2160: EGM08. Presented at the 2008 General Assembly of the European Geosciences Union, Vienna, Austria, 13–18 Apr 2008

Paul NK (1973) A method of evaluating the truncation error coefficients for geoidal height. Bull Geod 110:413–425
Prange L (2011) Global gravity field determination using the GPS measurements made onboard the low earth orbiting satellite CHAMP. Geodätisch-geophysikalische Arbeiten in der Schweiz, vol. 81. http://www.sgc.ethz.ch/sgc-volumes/sgk-81.pdf
Rapp R, Rummel R (1975) Methods for the computation of detailed geoids and their accuracy. Dept. of Geodetic Science and Surveying Report, No. 208. The Ohio State University, Columbus, Ohio
Sansó F (ed) (1997) Lecture notes. Int Geoid Service, DIIAR-Politecnico di Milano, Milan, Italy, Int School for the Determination and Use of the Geoid
Sansó F, Sideris MG (Eds.) (2013) Geoid determination theory and methods. Springer, Berlin. doi:10.1007/978-3-540-74700-0
Sjöberg LE (1981) Least squares combination of satellite and terrestrial data in physical geodesy. Ann Geophys 37(1):25–30
Sjöberg LE (1982) On the recovery of geopotential coefficients using satellite-to-satellite range-rate data on a sphere. Bull Geod 56:27–39
Sjöberg LE (1984a) Least-squares modification of Stokes' and Vening-Meinez' formula by accounting for truncation and potential coefficients errors. Manus Geod 9:209–229
Sjöberg LE (1984b) Least squares modification of Stokes' and Vening Meinesz' formulas by accounting for errors of truncation, potential coefficients and gravity data. Dept of Geodesy Rep No 27, Univ of Uppsala, Uppsala, Sweden
Sjöberg LE (1986) Comparison of some methods of modifying Stokes formula. Bollettino di Geodesia e Scienze Affini 46:229–248
Sjöberg LE (1991) Refined least-squares modification of Stokes' formula. Manus Geod 16:367–375
Sjöberg LE (2003a) A solution to the downward continuation effect on the geoid determined by Stokes' formula. J Geodesy 77:94–100
Sjöberg LE (2003b) A general model for modifying Stokes' formula and its least-squares solution. J Geod 77:459–464
Sjöberg LE (2005a) Discussion on the approximations made in the practical implementation of the remove-compute-restore technique in regional geoid modelling. J Geod 78:645–653
Sjöberg LE (2005b) A local least-squares modification of Stokes' formula. Stud Geophys Geod 49:23–30
Sjöberg LE (2011) Local least squares spectral filtering and combination by harmonic functions on the sphere. J Geod Sci 1(2011):355–360
Sjöberg LE (2014) On the topographic effects by Stokes' formula. J Geod Sci 4:130–135
Stokes GG (1849) On the variation of gravity on the surface of the earth. Trans Cambridge Phil Soc 8:672–695
Tapley BD, Flechtner F, Bettadpur SV, Watkins MM (2013) The status and future prospect for GRACE after the first decade, Eos Trans., Fall Meet. Suppl., Abstract G22A-01, 2013
Vanicek P, Kleusberg A (1987) The Canadian geoid ± Stokesian approach. Manuscr Geod 12:86–98
Vanicek P, Featherstone WE (1998) Performance of three types of Stokes's kernel in the combined solution for the geoid. J Geod 72:684–697
Vanicek P, Sjöberg LE (1991) Reformulation of Stokes's theory for higher than second-degree reference field and modification of integration kernels. J Geophys Res 96(B4):6529–6539
Vincent S, Marsh J (1974) Gravimetric global geoid. In: Veis G (ed) Proceedings of international symposium on the use of artificial satellites for geodesy and geodynamics. National Technical University, Athens, Greece
Wenzel H-G (1981) Zur Geoidbestimmung durch Kombination von Schwereanomalien und einem Kugelfunktionsmodell mit Hilfe von Integralformeln. Z Vermess 106:102–111

Wenzel H-G (1982) Geoid computation by least squares spectral combination using integration kernels. Proc IAG General Meet, Tokyo, pp 438–453
Wolff M (1969) Direct measurements of the Earth's gravitational potential using a satellite pair. J Geophys Res 74:5295–5300
Wong L, Gore R (1969) Accuracy of geoid heights from modified Stokes' kernels. Geophys J R Astr Soc 18:81–91

Chapter 5
Corrections in Geoid Determination

Abstract The original, as well as any type of modified Stokes' formula, requires a number of corrections as Stokes' integral allows no masses outside the sphere of integration. The corrections include direct topographic, atmospheric and ellipsoidal effects and a downward continuation (DWC) effect on the surface gravity anomaly to be applied prior to Stokes' integration. After integration, indirect effects are applied to the potential for restoration of masses as well as for corrections to the potential on the reference ellipsoid (rather than the sphere of integration). These are the classical corrections that are used more or less also in the modern remove-restore technique. The KTH method for geoid determination uses "additive corrections" to the preliminary geoid heights computed directly from the surface gravity anomalies. These corrections are therefore combinations of direct and indirect effects on potential/geoid height, implying several advantages. For example, the numerical solution to the DWC effect on the potential is much more stable than the corresponding effect on the gravity anomaly.

Keywords Atmospheric effects · Direct effects · Ellipsoidal effects · Indirect effects · Topographic effects · Topographic bias

5.1 Introduction

Stokes' formula (the original as well as the extended) is the solution to the boundary value problem of Eq. (3.23b) under the following conditions:

(a) There must be no masses outside the sphere of computation.
(b) The gravity anomalies are located on the sphere.

This implies that a number of corrections must be added to the gravity observations, and each type of correction is divided into a direct effect (e.g. the DITE) and an indirect effect (e.g. the PITE) (Sect. 3.4.4), where each effect means a correction. In contrast to the classical approach there is no secondary indirect topographic effect (SITE) on the gravity anomaly in the modern approach as (see Sjöberg 2014, 2015a):

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_5

$$DITE(\Delta g) = DITE(g) + SITE(g),$$

so that the SITE is already accounted for in the DITE (see Eq. (4.15a) and ensuing content, as well as Sect. 8.1).

The condition (a) above calls for the topographic and atmospheric corrections. Condition (b) requires that the surface gravity anomaly be downward continued to the sphere (downward continuation) prior to Stokes' integration, and the preliminary geoid height needs corrections by the indirect topographic and atmospheric effects after the integration. Usually the topographic effects are the largest corrections to be applied. The corrections are of the same order for geoid and quasigeoid determinations, except that the combined topographic effect in quasigeoid determination $\mathrm{DITE}(\zeta) + \mathrm{PITE}(\zeta)$ vanishes, implying that the combined effect cancels. As a result, formulas for quasigeoid determination (see Sect. 7.1.3) include no direct and indirect topographic corrections. [Nevertheless, they could still be practical in the computational steps for smoothing the data, a technique that is frequently applied in the RCR technique.] In addition, ellipsoidal effects will be considered.

Below we use the terminology correction and effect interchangeably. The corrections are generally of four kinds: *topographic, downward continuation (dwc), atmospheric* and *ellipsoidal corrections.*

The corrections in Sects. 5.2–5.5 mainly concern geoid determination, and, in Sect. 5.6, we briefly discuss the corrections for quasigeoid determination.

5.2 Topographic Corrections

As Stokes' formula does not allow external masses to the sphere of integration, the effect of these topographic masses are usually removed from the gravity anomaly (=*Direct Topographic Effect, DITE*, on the gravity anomaly) prior to Stokes' integration, and by integrating this effect in Stokes' formula, the DITE on the geoid height is obtained. Furthermore, after Stokes' integration, the co-geoid is corrected by the so-called *Primary Indirect Topographic Effect* (*PITE*) on the geoid height (corresponding to the restoration of topographic masses). Traditionally, in order to keep the effects small, the direct and indirect effects are reduced by some reduction method (e.g. Helmert's second method of condensation), but that will not be used here. In particular, as the KTH technique for geoid determination (or the Least Squares Modification of Stokes' Formula with Additive Corrections, LSMSA; see Chap. 6) adds the direct and indirect effects on the geoid heights to *a combined topographic effect*, the reduction for compensation potential cancels, and it is therefore not included in the presentation that follows.

We will assume that the topographic density (ρ) is constant, and we use the symbol μ for gravitational constant times ρ. (For more detailed derivations, see Sjöberg (2000, 2001.) In Sect. 5.2.8, we consider the effects of lateral topographic-density variations.

5.2.1 The Topographic Potential and Gravity Anomaly

The topographic potential is given by Newton's volume integral:

$$V^T(P) = \mu \iint_{\sigma} \int_{R}^{r_s} \frac{r^2}{l_P} dr d\sigma, \tag{5.1a}$$

where R and r_S are the radii of sea-level and topographic height and

$$l_P = \sqrt{r_P^2 + r^2 - 2rr_P t} \quad \text{with} \quad t = \cos\psi. \tag{5.1b}$$

By expanding $1/l_P$ in an external type series (cf. Eqs. 2.54a and b), $V^T(P)$ becomes

$$V^T(P) = \mu \sum_{n=0}^{\infty} \iint_{\sigma} \int_{R}^{r_S} \frac{r^{n+2}}{r_P^{n+1}} dr P_n(t) d\sigma = \mu R^2 \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+1} a_n(P), \tag{5.2a}$$

where:

$$a_n(P) = \frac{1}{n+3} \iint_{\sigma} \left[(1 + H/R)^{n+3} - 1\right] P_n(t) d\sigma. \tag{5.2b}$$

By expanding $(1 + H/R)^{n+3}$ in a Taylor series of H/R and omitting terms of order four and higher, we obtain:

$$a_n \approx \mu \iint_{\sigma} \left[\frac{H}{R} + \frac{n+2}{2}\left(\frac{H}{R}\right)^2 + \frac{(n+2)(n+1)}{6}\left(\frac{H}{R}\right)^3\right] P_n(t) d\sigma. \tag{5.3}$$

Using the boundary condition in the spherical approximation, Eq. (3.23b), we get also the topographic gravity anomaly (see also 4.15c):

$$\Delta g^T(P) = -\frac{\partial V^T(P)}{\partial r_P} - 2\frac{V^T(P)}{r_P} = \mu R \sum_{n=0}^{\infty} \left(\frac{R}{r_P}\right)^{n+2} (n-1) a_n(P). \tag{5.4}$$

Of particular interest here is the negative of the downward continued topographic gravity anomaly, i.e. the *direct gravity anomaly effect at sea level*:

$$\delta\Delta g^*_{dir} = -\left(\Delta g^T(Q)\right)_{r_Q = R} = \sum_{n=0}^{\infty} \left(\delta\Delta g^*_{dir}\right)_n. \tag{5.5}$$

By inserting it into Stokes' formula, one obtains the *direct topographic effect on the geoid height*:

$$\delta N^T_{dir}(P) = -\frac{R}{4\pi\gamma}\iint_\sigma S(\psi)\delta\Delta g^*_{dir}d\sigma. \tag{5.6}$$

By considering the spectral form of Stokes' function and the orthogonality for it and the Laplace series (5.5), one finally obtains:

$$\delta N^T_{dir}(P) = \frac{R}{\gamma}\sum_{n=2}^{\infty}\frac{(\delta\Delta g^*_{dir})_n}{n-1} = -\frac{R^2\mu}{\gamma}\sum_{n=2}^{\infty}a_n(P). \tag{5.7}$$

Note. Equations (5.7) and (5.3) show that the direct effect on the geoid contains all orders of H/R. If the direct effect were applied with Helmert's method of condensation, the difference would have been that the term of order H/R would vanish.

5.2.2 The Indirect Effect on the Geoid

By expanding $1/l_P$ as an internal type series (see Sect. 2.4) with $r_P = R$, one obtains the topographic potential at the geoid as the Laplace series (with exclusion of terms of degrees 0 and 1; see the next section):

$$V^T_g(P) = \mu R^2\sum_{n=2}^{\infty}b_n(P), \tag{5.8a}$$

where:

$$b_n(P) = \frac{\mu}{n-2}\iint_\sigma\left[1-(1+H/R)^{-n+2}\right]P_n(t)d\sigma. \tag{5.8b}$$

A Taylor expansion of (5.8b) yields:

$$b_n(P) \approx \mu\iint_\sigma\left[\frac{H}{R}-\frac{n-1}{2}\left(\frac{H}{R}\right)^2+\frac{(n-1)n}{6}\left(\frac{H}{R}\right)^3\right]P_n(t)d\sigma. \tag{5.9}$$

Note. If Helmert's condensation potential were applied to reduce the indirect effect, the first term under the integral of Eq. (5.9) would vanish.

5.2.3 The Combined Effect on the Geoid

The combined topographic effect on the geoid height is the sum of the direct and indirect effects:

$$\delta N^T_{comb}(P) = \delta N^T_{dir}(P) + \delta N^T_I(P) = \frac{\mu R^2}{\gamma}\sum_{n=2}^{\infty}[b_n(P) - a_n(P)]. \tag{5.10a}$$

and the Taylor series (5.3) and (5.9) yield:

$$\delta N^T_{comb}(P) \approx -\frac{\mu}{\gamma}\sum_{n=2}^{\infty}\iint_{\sigma}(2n+1)\left[\frac{H^2}{2}+\frac{H^3}{3R}\right]P_n(t)d\sigma = -\frac{2\pi\mu}{\gamma}\left(\widetilde{H}^2+\frac{2\widetilde{H}^3}{3R}\right), \tag{5.10b}$$

where $\widetilde{H}^{\nu}$ is the νth power of the topographic height with first- and second-degree terms excluded.

Finally, we mention that, if one computes the combined effect directly from the direct and indirect effects without the above Taylor series for topographic height, one obtains the Newton integral (Sjöberg 2000)

$$\delta N^T_{comb}(P) = \frac{\mu}{R\gamma}\sum_{n=2}^{\infty}\iint_{\sigma}\int_{R}^{r_s}\left[\left(\frac{R}{r}\right)^{n+1}-\left(\frac{r}{R}\right)^{n}\right]r^2 dr P_n(t)d\sigma. \tag{5.11}$$

It is obvious that this series in Legendre polynomials could hardly converge.

5.2.4 Zero- and First-Degree Effects

In the above derivation of the combined topographic effect, terms of orders 0 and 1 were not included. This was the case, as Stokes' formula cancels these effects in the direct topographic effect. Nevertheless, a correct geoid determination with Stokes' formula needs also contributions from these terms (despite that, the external gravity field is defined in such a way that such terms vanish in the potential) and gravity anomaly. More precisely: the definition that the external gravity field contains no such harmonics does not warrant that they vanish at the geoid, inside the topography. This will now be investigated.

From Eq. (5.2a), we obtain (Sjöberg 2001):

$$\left(\delta N^T_{dir}\right)_{0,1} = -\frac{\mu R^2}{\gamma}[a_0(P) + a_1(P)], \tag{5.12}$$

which, after inserting (5.2b) and expanding the powers and considering the addition theorem of spherical harmonics, Eq. (2.49), leads to:

$$\left(\delta N^T_{dir}\right)_{0,1} = -\frac{4\pi\mu}{\gamma}\left[H_0^2 + \frac{H_0^3}{3R} + H_1^2 + \frac{H_1^3}{3R} + \frac{H_1^4}{12R^2}\right]. \tag{5.13}$$

Similarly, the indirect geoid effects of degrees 0 and 1 is given by Eqs. (5.8a, b):

$$\left(\delta N^T_I\right)_{0,1} = \frac{\left(V_g^T\right)_0 + \left(V_g^T\right)_1}{\gamma} = \frac{2\pi\mu}{\gamma} H_0^2, \tag{5.14}$$

and, by adding direct and indirect effects, the combined effect becomes:

$$\left(\delta N^T_{comb}\right)_{0,1} = -\frac{2\pi\mu}{\gamma}\left[H_0^2 + \frac{2H_1^2}{3R} + \frac{H_1^4}{6R^2}\right]. \tag{5.15}$$

If we add Eqs. (5.10b) and (5.15) (truncated to power $(H/R)^3$), we finally obtain:

$$\delta N^T_{comb}(P) \approx -\frac{2\pi\mu}{\gamma}\left[H^2(P) + \frac{2}{3}\frac{H^3(P)}{R}\right]. \tag{5.16}$$

At this point, it should be emphasized that the series with harmonics $a_n(P)$ in the strict formula (5.10a) and the first term under the integral in (5.11) are not likely to converge [in contrast to the approximation (5.16)]. As discussed in Sjöberg (2007, 2009a, b, c; see also next section), Eq. (5.16) corresponds (exactly) to the negative of the so-called *topographic potential bias*, and, in this view, this equation is the strict combined effect on the geoid height. It is interesting that this deviation of the spherical harmonic series from correct downward continuation to the geoid within the masses shows up already in the first-degree term (the last term of Eq. 5.15, which is not present in 5.16). Moreover, Eq. (5.16) proves also, in contrast to common belief, that the geoid contains zero- and first-degree harmonics.

Importantly, the combined topographic effect does not include any terrain effect in opposition to the direct and indirect topographic effects. This fact very significantly simplifies the computation of the total topographic effect.

The combined topographic effect of Eq. (5.16) is very substantial, reaching 9.8 m for Mt. Everest. However, much of this effect is counteracted by the downward continuation effect, discussed in Sect. 5.3.

5.2.5 The Topographic Bias by a Strict Formulation

The method of harmonic, analytical continuation discussed so far can be applied only to determine the gravity related quantities on, and exterior to, the Earth's surface, while its application inside the Earth will lead to fictitious or biased results, as the potential is analytically continued into the space where it is not harmonic. For example, this will occur if the geoid is estimated in continental areas. In such a case, the exterior type of disturbing potential analytically continued down to the geoid (T^*) will disagree with the true disturbing potential at the geoid $\left(T_g\right)$ by the potential bias

$$dT_{bias} = T^* - T_g = \left(T_g^{NT} + \left(V^T\right)^*\right) - \left(T_g^{NT} + V_g^T\right) = \left(V^T\right)^* - V_g^T, \tag{5.17}$$

so that the geoid height determined by Bruns' formula gets the bias

$$dN_{bias} = \frac{\left(V^T\right)^* - V_g^T}{\gamma_0}. \tag{5.18}$$

The question is whether this bias can be determined to correct the solution. The answer is in the affirmative, as shown in various ways by Sjöberg (2007, 2009a, b, c). Here we mainly follow Sjöberg (2007) to demonstrate the idea.

Let us decompose the topographic potential V^T into those of a Bouguer shell (V^B) and the terrain (V^t):

$$V^T = V^B + V^t. \tag{5.19}$$

The terrain is all topographic mass outside the Bouguer shell (excess masses), as well as the negative (deficient) masses located inside the Bouguer shell (see Fig. 5.1).

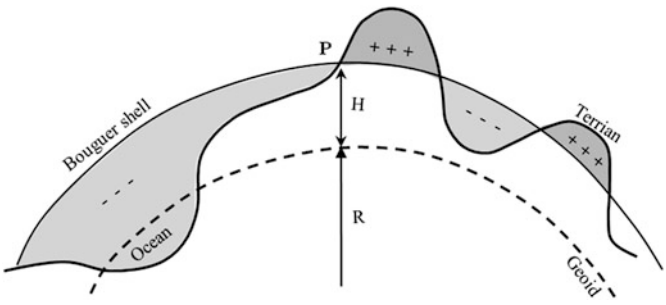

Fig. 5.1 Schematic diagram of the terrain (*shaded*) with excess and deficient masses outside and inside the Bouguer shell, respectively

For simplicity, we assume that the topographic density (ρ) is constant, implying that the Bouguer shell and terrain potentials can be written by the Newton integrals

$$V_P^B = \mu \iint_{\sigma} \int_{R}^{\Re} \frac{r^2 dr}{l_P} d\sigma = 2\pi\mu \begin{cases} \frac{2}{3r_P}(\Re^3 - R^3), & if \quad r_P \geq \Re \\ \Re^2 - \frac{2R^3}{3r_P} - \frac{r_P^2}{3}, & if \quad r_P \leq \Re \end{cases} \tag{5.20a}$$

and

$$V^t = \mu \iint_{\sigma} \int_{\Re}^{r_s} \frac{r^2 dr}{l_P} d\sigma, \tag{5.20b}$$

where μ is the gravitational constant times the topographic density, r_P and $\Re$ are the geocentric radii of the computation point P and the topographic surface along the radius through P, respectively. Hence, $\Re = r_S(P)$, where r_S is the surface radius at the integration point.

Then $(V^B)^*$ is given by the exterior type of solution $(r_P \geq \Re)$ in Eq. (5.20a) whence r_P is forced to R (analytical continuation of the exterior solution). By introducing the orthometric height of the topography H, it follows with $\Re = R + H$ that the contribution from the Bouguer shell to the topographic bias of the potential at the geoid becomes:

$$V_{bias}^B = (V^B)^* - V_g^B = 2\pi\mu\left(H^2 + \frac{2H^3}{3R}\right). \tag{5.21}$$

It remains to determine the bias of the terrain potential V^t. However, as discussed in Sjöberg (2007), this potential does not contribute to the bias. Here we limit this discussion to consider the Laplace equation for the topographic potential, having in mind that there will be no topographic bias in the analytical continuation of the potential in free air, where the potential is harmonic. Inside the topography one obtains:

$$\Delta V^T = \Delta V^B + \Delta V^t, \tag{5.22}$$

and here both V^T and V^B obey Poisson's formula:

$$\Delta V^T = \Delta V^B = -4\pi\rho, \tag{5.23}$$

so that (5.22) yields $\Delta V^t = 0$, which means that the terrain potential is harmonic and therefore does not contribute to the potential bias, and the total potential bias is therefore given by Eq. (5.21).

For any radius r inside the topography the topographic potential bias can be written as:

$$b(r) = \frac{2\pi\mu}{3}\left(\frac{2\Re^3}{r} - 3\Re^2 + r^2\right); \quad R \le r < \Re, \tag{5.24}$$

yielding:

$$\Delta b(r) = \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right) b(r) = 4\pi\mu; \quad R \le r < \Re \tag{5.25}$$

and the bias and its radial derivative vanish at the topographic surface.

As T^* is harmonic all the way down to sea level, the corrected analytically continued disturbing potential obeys Poisson's differential equation:

$$\Delta(T^* - b(r)) = -4\pi\mu; \quad R \le r < \Re. \tag{5.26}$$

In Sjöberg (2007, 2009c), Eq. (5.24) is generalized to an arbitrary density distribution within the topography.

Vermeer (2008) objected to this result and claimed that Eq. (5.21) is only an approximation to the bias when given by the potential of the Bouguer shell. However, Sjöberg (2008, 2009a, b, c) confirmed in alternative proofs that Eq. (5.21) contains the total topographic bias. This implies that Bjerhammar's method and collocation can also be used to determine the geoid height after considering the bias. Another consequence of this result is that the topographic effect in geoid determination is independent on the terrain effect (Sjöberg 2009a), which is taken advantage of in the KTH method of geoid determination (See Chap. 6.). Finally, our claim that Eq. (5.21) is the total topographic bias is supported by the following circumstances for any point P located on the Earth's surface at the geoid level with no topographic masses along the vertical above P:

The geoid and quasigeoid heights are the same.

As the quasigeoid heights have no topographic bias, this holds as well for the geoid height.

The topographic bias (i.e. the potential given by the Bouguer shell) determined by Eq. (5.21) correctly vanishes, independent on the slope of the terrain, and therefore: There is no terrain effect.

Equation (5.21) is also supported by the following examples.

Example 5.1 If the geoid is located above the Earth's surface, Eq. (5.21) correctly suggests that the topographic correction vanishes.

Proof The determination of the disturbing potential on the Earth's surface (cf. the height anomaly) or outside the surface (this example) can be determined by the

extended Stokes formula by analytical continuation, which need no topographic correction. See Sects. 7.1.2 and 7.1.3.

Example 5.2 Consider an Earth model in the shape of a sphere of radius $R + H$, but with a hole in the form of a spherical cap of height H and geocentric angle v. Then there is no topographic bias along the axis of the hole.

Proof The potential at geocentric radius r_P of a spherical (Bouguer) shell of density ρ and thickness H can be written (Sjöberg 2007):

$$V^{shell}(r_P) = 2\pi\mu \begin{cases} \frac{2\left(r_s^3 - R^3\right)}{3r_P} & if \quad r_P \geq r_s = R + H \\ \left(r_s^2 - R^2\right) & if \quad r_P \leq r_s \end{cases}, \tag{5.27}$$

where $\mu = G\rho$, G being the gravitational constant and R the inner radius of the shell. The topographic bias of the shell is obtained by taking the difference between the exterior potential downward continued to sea level with $r_P = R$ and the internal potential at the same radius:

$$bias^{shell} = \left(V^{shell}\right)^* - V^{shell}_{r_P=R} = 2\pi\mu\left(H^2 + \frac{2H^3}{3R}\right). \tag{5.28}$$

Similarly, the potential along the symmetry axis of a spherical cap of geocentric angle v and height H on top of a sphere of radius R becomes (Sjöberg 2007):

$$V_P^{cap} = 2\pi\mu \int_R^{r_s} \int_0^v \frac{r^2 dr \, \sin\psi d\psi}{l_P} = \frac{2\pi\mu}{r_P} \int_R^{r_s} (l_P - |r_P - r|) r dr \tag{5.29a}$$

with:

$$V_P^{cap} = \frac{2\pi\mu}{r_P}\left[I(r_P) - r_P \frac{2RH + H^2}{2} + \frac{r_s^3 - R^3}{3}\right], \quad \text{if} \quad r_P \geq r_s \tag{5.29b}$$

and

$$V_g^{cap} = \frac{2\pi\mu}{R}\left[I(R) - \frac{RH^2}{2} - \frac{H^3}{3}\right], \quad \text{if} \quad r_P = R, \tag{5.29c}$$

where:

$$I(r_P) = \int_R^{r_s} l_P r dr. \tag{5.29d}$$

Hence, the potential bias along the axis of the cap becomes:

$$bias^{cap} = \left(V_{r_P}^{cap}\right)^{*} - V_{r_P=R}^{cap} - = 2\pi\mu\left[-\frac{2RH+H^2}{2} + \frac{r_s^3 - R^3}{3R} + \frac{H^2}{2} + \frac{H^3}{3R}\right]$$
$$= 2\pi\mu\left(H^2 + \frac{2H^3}{3R}\right) \tag{5.30}$$

which is exactly the same as for the shell.

Consider now a spherical shell with a hole in the form of a spherical shaped cap with the above dimensions. Then the potential bias along the symmetry axis becomes

$$bias^{shell} - bias^{cap} = 0, \tag{5.31}$$

i.e. there is no bias for this model.

5.2.6 The EGM Analytical Continuation Error (EACE)

In this section, we will see that the problem studied by Sjöberg (1977) and Vermeer (2008), which we name the *EGM analytical continuation error* (EACE), is different from the topographic bias in Stokes' formula. Here we closely follow the presentation of Sjöberg (2015b).

– *Application of the EGM at Surface Level*

Equation (5.21) holds only if the analytical continuation exists, which is not necessarily the case for an external type series of an EGM. On one hand, in some cases the series could converge also inside topographic masses. One such example is the external series of the potential of a homogeneous oblate ellipsoid, which converges all the way down to the radius $r > ae =$ the radius of the focal points. On the other hand, for the irregular mass distribution of the Earth, it is doubtful whether the external type series converges in the strict sense inside the Brillouin sphere bounding all mass of the Earth. In particular, we should not expect that such a series (see Eq. (5.35) below) converges inside the topographic masses. However, in any case, we may assume that the series is asymptotically convergent (see Moritz 1980, pp. 413–414; Moritz 2003) outside the topographic masses, and, as the series representation by an EGM is always finite, we do not have to worry about the divergence, but the series is "practically" convergent. Alternatively, one may represent the dwc by a Taylor series:

$$\left(V_P^T\right)^{*} = \sum_{k=0}^{\infty} \frac{(-H_P)^k}{k!} \frac{\partial^k V_P^T}{\partial r_P^k}, \tag{5.32}$$

where point P is located on or outside the Earth's surface and the derivatives are determined from the external type harmonic series. The Taylor series is not likely to converge inside the topographic masses, but, as in practice such a series s always limited to a finite degree of expansion, it will be "practically convergent".

Below we consider the problem of applying the EGM directly at sea level, which is not the same as in Eq. (5.32).

Application of the EGM at Sea Level.

Disregarding the mass of the atmosphere, the Earth's total potential V can be decomposed into the no-topography potential V^{NT} and the topographic potential V^T:

$$V = V^{NT} + V^T. \tag{5.33}$$

The Newton integral for the latter component reads:

$$V^T = \iint_{\sigma} \int_{R}^{r_s} \mu \frac{r^2 dr}{l} d\sigma, \tag{5.34}$$

where l is the distance between the computation and integration points. The above potentials can be expanded into external types of spherical harmonic series at any point with geocentric radius $r > r_b$, where r_b is the bounding radius of the Earth's topography:

$$\begin{aligned} V(r,\theta,\lambda) &= \sum_{n=0}^{\infty} \left(\frac{r_b}{r}\right)^{n+1} \sum_{m=-n}^{n} V_{nm}^{e} Y_{nm}(\theta,\lambda) \\ &= \sum_{n=0}^{\infty} \left(\frac{r_b}{r}\right)^{n+1} \sum_{m=-n}^{n} \left(V_{nm}^{NT} + V_{nm}^{T,e}\right) Y_{nm}(\theta,\lambda), \end{aligned} \tag{5.35}$$

where V_{nm}^{NT} and $V_{nm}^{T,e}$ are the harmonic coefficients for the series of potentials of Eq. (5.33). The series for the no-topography potential is actually convergent all the way down to sea level, while the series of the topographic potential is likely to be divergent for $r < r_b$, or at least so for any point located inside the topography.

In practice, the gravitational potential can be represented by such a series limited to some maximum degree M:

$$\begin{aligned} \widehat{V}(r,\theta,\lambda) &= \sum_{n=0}^{M} \left(\frac{r_b}{r}\right)^{n+1} \sum_{m=-n}^{n} V_{nm}^{e} Y_{nm}(\theta,\lambda) \\ &= \sum_{n=0}^{M} \left(\frac{r_b}{r}\right)^{n+1} \sum_{m=-n}^{n} \left(V_{nm}^{NT} + V_{nm}^{T,e}\right) Y_{nm}(\theta,\lambda), \end{aligned} \tag{5.36}$$

and the correct series for the topographic potential at sea level is an internal type series:

$$V(R,\theta,\lambda)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}V_{nm}^{i}Y_{nm}(\theta,\lambda)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[\left(\frac{r_b}{R}\right)^{n+1}V_{nm}^{NT}+V_{nm}^{T,i}\right]Y_{nm}(\theta,\lambda), \tag{5.37}$$

The harmonic coefficients $V_{nm}^{T,e}$ and $V_{nm}^{T,i}$ are given explicitly by Eqs. (2.55b and 2.55c) for $r_P=R$. This means that the error for the external type harmonic series $\widehat{V}(R,\theta,\lambda)$, complete to degree M, applied at sea level with radius R, consists of the sum of *the EGM analytically continuation error* (*EACE*) (or commission error) and the *truncation error* (*ome;* omission error):

$$\varepsilon_V(\theta,\lambda)=\widehat{V}(R,\theta,\lambda)-V(R,\theta,\lambda)=EACE_M(\theta,\lambda)+ome(\theta,\lambda), \tag{5.38a}$$

where:

$$EACE_M(\theta,\lambda)=\sum_{n=0}^{M}\sum_{m=-n}^{n}\left(V_{nm}^{T,e}-V_{nm}^{T,i}\right)Y_{nm}(\theta,\lambda) \tag{5.38b}$$

and

$$ome(\theta,\lambda)=-\sum_{n=M+1}^{\infty}\sum_{m=-n}^{n}V_{nm}^{T,i}Y_{nm}(\theta,\lambda) \tag{5.38c}$$

are the commission and omission bias, respectively. [Ågren (2004a) denoted the dwc error the analytical continuation bias.] Assuming that the topographic density changes only laterally, it can be denoted $\mu(\theta,\lambda)$ and the *EACE* can be written (Sjöberg 1977, p. 16) and Ågren (2004a):

$$EACE_M(\theta,\lambda)=\sum_{n=0}^{M}\sum_{m=-n}^{n}A_{nm}Y_{nm}(\theta,\lambda) \tag{5.39a}$$

where:

$$A_{nm}=\frac{R^2}{2n+1}\iint_{\sigma}\mu(\theta,\lambda)I_n(R,r_s)Y_{nm}(\theta,\lambda)d\sigma, \tag{5.39b}$$

with:

$$I_n(R,r_s)=\begin{cases}\frac{(r_s/R)^{n+3}-1}{n+3}+\frac{(r_s/R)^{-(n-2)}-1}{n-2} & if \quad n\neq 2\\ \frac{(r_s/R)^5-1}{5}-\ln(r_s/R) & if \quad n=2\end{cases}. \tag{5.39c}$$

It is obvious from Eqs. (5.39a–5.39c) that the $EACE_M$ diverges when n approaches infinity. However, in practices, the harmonic series is always finite.

Considering that $r_s = R + H$ and the Taylor series

$$(r_s/R)^{n+3} = (1 + H/R)^{n+3} = \sum_{k=0}^{\infty} \binom{n+3}{k} \left(\frac{H}{R}\right)^k = 1 + (n+3) \sum_{k=1}^{\infty} \frac{1}{k} \binom{n+2}{k-1} \left(\frac{H}{R}\right)^k \tag{5.40a}$$

and

$$(r_s/R)^{-(n-2)} = (1 + H/R)^{-(n-2)} = \sum_{k=0}^{\infty} \binom{-(n-2)}{k} \left(\frac{H}{R}\right)^k = 1 - (n-2) \sum_{k=1}^{\infty} \frac{1}{k} \binom{-n+1}{k-1} \left(\frac{H}{R}\right)^k, \tag{5.40b}$$

one obtains also:

$$I_n(R, r_s) = \sum_{k=1}^{\infty} \frac{1}{k} \left[\binom{n+2}{k-1} - \binom{-n+1}{k-1} \right] \left(\frac{H}{R}\right)^k = (2n+1) R^{-2} \left[H^2 + \frac{2}{3} \frac{H^3}{R} \right] + \sum_{k=4}^{\infty} \frac{1}{k} \left[\binom{n+2}{k-1} - \binom{-n+1}{k-1} \right] \left(\frac{H}{R}\right)^k \tag{5.41}$$

Assuming now that the topographic density is constant, and by inserting Eq. (5.41) into Eq. (5.39b), the *EACE* limited to degree *M*, can be written (Sjöberg 1977; Martinec 1998a; Sect. 7.4; Ågren 2004a)

$$EACE_M = \sum_{n=0}^{M} (ACE)_n = 2\pi\mu \left[H_M^2 + \frac{2}{3} \frac{H_M^3}{R} \right] + \sum_{n=0}^{M} C_n, \tag{5.42a}$$

where

$$C_n = C_n(\theta, \lambda) = \frac{4\pi R^2 \mu}{2n+1} \sum_{k=4}^{\infty} \frac{1}{k} \left[\binom{n+2}{k-1} - \binom{-n+1}{k-1} \right] \frac{(H^k)_n}{R^k} \tag{5.42b}$$

and

$$H_M^i = \sum_{n=0}^{M} (H^i)_n; \quad \mathrm{i} = 2, 3. \tag{5.42c}$$

Here $(H^i)_n$ are Laplace harmonics of H^i. It is obvious that the EACE of the external type harmonic series is not only the topographic bias of Eq. (5.21), but it also includes an additional contribution due to the divergence of the series when applied at sea level.

Note that the sea-level radius is approximated by the mean Earth-sphere radius R and not the actual radius of the reference ellipsoid. As already mentioned in the introduction, the correction from the sphere to the reference ellipsoid is best treated by a separate ellipsoidal correction.

Example 5.3 Reconsider the Earth model of Example 5.2. Here we represent the potential of the shell with a hole by an harmonic series limited to upper degree n_0. The potential of the shell is limited to the zero-degree harmonics

$$V^{shell}(r_P) = 2\pi\mu \begin{cases} \frac{2}{3r_P} & if \quad r_P \geq r_s = R + H \\ (r_s^2 - R^2) & if \quad r_P \leq r_s \end{cases}, \tag{5.43}$$

and the potential along the axis of the spherical cap represented as a harmonic series becomes (see Sjöberg 2015b):

$$I^e(r_P) = 2\pi\mu \sum_{n=0}^{n0} \frac{r_s^{n+3} - R^{n+3}}{(n+3) r_P^{n+1}} J_n(v) \quad \text{if} \quad r_P \geq r_s \tag{5.44a}$$

and

$$I^i(r_P) = 2\pi\mu \sum_{\substack{n=0 \\ n \neq \text{`2}}}^{n0} \left(\frac{r_P^n}{R^{n-2}} - \frac{r_P^n}{r_s^{n-2}} \right) \frac{J_n(v)}{n-2} + 2\pi\mu r_P^2 \ln \frac{r_s}{R}, \tag{5.44b}$$

where:

$$\begin{aligned} J_n(v) &= \int_o^v P_n(\cos\psi) \sin\psi d\psi \\ &= \begin{cases} 1 - \cos\psi & n = 0 \\ [P_{n-1}(\cos v) - P_{n+1}(\cos v)]/(2n+1) & \text{otherwise} \end{cases}. \end{aligned} \tag{5.44c}$$

Hence, for $r_P \to R$ the topographic bias becomes:

$$\begin{aligned} V_{bias} = I^e(R) - I^i(R) = 2\pi\mu R^2 \Bigg[\sum_{\substack{n=0 \\ n \neq 2}}^{n0} J_n(v) \Big\{ \tfrac{(1+H/R)^{n+3}-1}{n+3} \\ - \tfrac{1}{n-2} \left(1 - (1+H/R)^{-n+2} \right) + J_2(v) \ln(1+H/R) \Big\} \Bigg] \end{aligned} \tag{5.45}$$

It is obvious that Eq. (5.45) diverges as $n \to \infty$.

5.2.7 The Topographic Bias in the Modified Stokes' Formula

– *Modification with Full-Gravity Anomaly*

Consider the unbiased type of estimator for the modified Stokes' formula, augmented by zero- and first-degree disturbing potential harmonics (T_0 and T_1) to the absolute geoid height (Sjöberg 2003b):

$$\widetilde{N} = \frac{T_0^* + T_1^*}{\gamma} + \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\Delta g^* d\sigma + \sum_{n=2}^{M}\left(s_n + Q_n^L\right)\frac{n-1}{2}N_n^{EGM}; \quad L \geq M, \tag{5.46a}$$

where L is the maximum degree of modification of Stokes' function, σ_0 is a spherical cap of geocentric angle ψ_0 around the computation point, $c = R/(2\gamma)$, Δg_n^{EGM} is the Laplace harmonic of Δg determined from the EGM complete to degree M and downward continued to sea level of radius R, $Q_n^L = Q_n^L(\psi_0)$ are the modified Molodensky truncation coefficients and:

$$S^L(\psi) = S(\psi) - \sum_{k=2}^{L}\frac{2k+1}{2}s_k P_k(\cos\psi) \tag{5.46b}$$

is the modified Stokes function. Here $P_k(\cos\psi)$ is the Legendre's polynomial of degree k, and s_k are the (arbitrary) modification parameters.

In Eqs. (5.46a), Δg^* and N_n^{EGM} are biased in continental areas (as they are determined by harmonic downward continuation). Hence, the above estimator needs a correction for the biases, which can be expressed by:

$$\delta N_{corr}^T = \delta N_{corr,1}^T + \delta N_{corr,2}^T, \tag{5.47}$$

where $\delta N_{corr,1}^T$ is the correction for topographic bias in the Stokes' integral, and $\delta N_{corr,2}^T$ is the correction for the EACE in the harmonic series. The integral can be written in spectral form as:

$$\frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\Delta g^* d\sigma = c\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - Q_n^L - s_n^*\right)(\Delta g_n)^*, \tag{5.48}$$

which [after considering also the boundary condition of physical geodesy; (Heiskanen and Moritz 1967, p. 89)] needs the following correction for the topographic bias:

$$\delta N^T_{corr,1} = -\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - Q^L_n - s^*_n\right)_n \frac{(n-1)}{2}\left(\delta N^T_{bias}\right)_n \tag{5.49a}$$

where:

$$s^*_n = \begin{cases} s_n & if \quad n \le L \\ 0 & otherwise \end{cases}. \tag{5.49b}$$

The second term in Eq. (5.47) becomes:

$$\delta N^T_{corr,2} = -\sum_{n=2}^{M}\left(Q^L_n + s_n\right)\frac{n-1}{2\gamma} EACE_n, \tag{5.50}$$

so that the total topographic correction can be written as:

$$\delta N^T_{corr} = -\delta N^T_{bias} - \frac{1}{\gamma}\sum_{n=0}^{M}\left(Q^L_n + s^*_n\right)\frac{n-1}{2}C_n + \sum_{n=M+1}^{\infty}\left[\frac{n-1}{2}\left(Q^L_n + s^*_n\right)\right]\left(\delta N^T_{bias}\right)_n. \tag{5.51}$$

where C_n was given in Eq. (5.42b), and the *bias* term is the same as for the original Stokes' formula. We notice that for $M = 0$ the corrections is that for Stokes' modified integral truncated to a cap σ_0, and for $\sigma_0 = 0$ it follows that $Q^L_n = 2/(n-1) - s^*_n$.

– *Modification with Reduced-Gravity Anomaly*

As shown by Sjöberg (2005), Eq. (5.46a) can be re-written for a reduced gravity anomaly ("higher order reference field") as:

$$\widetilde{N} = \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\left(\Delta g^M\right)^* d\sigma + \sum_{n=0}^{M} N^{EGM}_n, \tag{5.52a}$$

where (when disregarding observation errors):

$$\Delta g^M = \Delta g - \sum_{n=2}^{M}\Delta g^{EGM}_n = \sum_{n=M+1}^{\infty}\Delta g_n. \tag{5.52b}$$

In Sjöberg (2005), no analytical continuation effects were considered in showing that Eqs. (5.46a) and (5.52a) are equal. As these effects differ for Stokes' integral and the harmonic series, the total topographic errors in Eqs. (5.46a) and (5.52a) due to analytical continuation will also differ.

In Eq. (5.52a), $(\Delta g^M)^*$ and N_n are in error in continental areas (as they are determined by harmonic downward continuation), requesting a correction, which in this case can be expressed as:

$$\delta N^T_{corr} = \delta N^T_{corr,1} + \delta N^T_{corr,2}, \tag{5.53}$$

where $\delta N^T_{corr,1}$ is the correction for topographic bias in the Stokes' integral, and $\delta N^T_{corr,2}$ is the correction for the ACE in the harmonic series. The integral can be written in spectral form as:

$$\frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\left(\Delta g^M\right)^* d\sigma = c\sum_{n=M+1}^{\infty}\left(\frac{2}{n-1} - Q^L_n - s^*_n\right)(\Delta g_n)^*, \tag{5.54}$$

which needs the following correction for the analytical continuation:

$$\begin{aligned}\delta N^T_{corr,1} &= -\sum_{n=M+1}^{\infty}\left(\frac{2}{n-1} - Q^L_n - s^*_n\right)_n \frac{(n-1)}{2}\left(\delta N^T_{bias}\right)_n \\ &= \sum_{n=M+1}^{\infty}\left[-1 + \frac{n-1}{2}\left(Q^L_n + s^*_n\right)\right]\left(\delta N^T_{bias}\right)_n.\end{aligned} \tag{5.55}$$

The second term in Eq. (5.53) becomes:

$$\delta N^T_{corr,2} = -\frac{EACE_M}{\gamma}, \tag{5.56}$$

so that the total correction becomes:

$$\delta N^T_{corr} = -\delta N^T_{bias} - \frac{1}{\gamma}\sum_{n=0}^{M} C_n + \sum_{n=M+1}^{\infty}\frac{n-1}{2}\left(Q^L_n + s^*_n\right)\left(\delta N^T_{bias}\right)_n. \tag{5.57}$$

where C_n was given in Eq. (5.42b). The *bias* term is the same as for the original Stokes' formula. We notice that for $M = 0$ the correction is that for a Stokes' modified integral truncated to a cap σ_0, and for $\sigma_0 = 0$ it follows that $Q^L_n = 2/(n-1) - s^*_n$ and Eq. (5.57) equals $-(EACE)_M/\gamma$, i.e. the correction for the spherical harmonic representation of Eq. (5.42a).

5.2.8 *Lateral Topographic Density Variations*

If the actual topographic density differs laterally from normal density by $\Delta\mu$, it follows directly from Eq. (5.21) that the combined effect on the geoid height differs from its normal effect by (Sjöberg 2004a):

$$\delta N^T_{\Delta\mu} = -\frac{2\pi\Delta\mu}{\gamma}\left[H^2 + \frac{2}{3}\frac{H^3}{R}\right]. \tag{5.58}$$

It is obvious that this topographic correction is purely local, and there is no need to consider terrain effects. As a matter of fact, the terrain effect cancel when adding the direct and indirect geoid effect, and the remainder is the Bouguer shell effect (see, e.g. Sjöberg 2009a).

Note. If $H < 0$ (Earth's surface below sea level), all the topographic effects on the geoid estimate cancel.

5.3 The Downward Continuation Correction

The downward continuation (dwc) effect is the correction to the geoid height due to the fact that gravity data, located at the surface of the Earth, must be reduced to sea level (approximated by a sphere) to be consistent with Stokes' formula.

5.3.1 The Dwc Effect on the Original Stokes' Formula

The dwc effect (or correction) for the geoid height, as determined approximately by the original Stokes' formula with surface gravity data, is given by:

$$dN_{dwc} = \frac{R}{4\pi\gamma_0}\iint_\sigma S(\psi)(\Delta g^* - \Delta g)d\sigma, \tag{5.59}$$

where the symbols are the same as above. As discussed in Sjöberg (2003b), the first part of Eq. (5.59) is nothing but the downward continued disturbing potential T^* divided by normal gravity, and the latter part is the negative of an approximation to the disturbing potential given at the surface (T), again divided by normal gravity, i.e.

$$dN_{dwc} = \frac{T^* - (T)}{\gamma_0}. \tag{5.60}$$

Denoting the computational point by P and its geocentric radius by r_P, and introducing the auxiliary potential

$$\widetilde{T} = \frac{R}{4\pi\gamma_0}\iint_\sigma S(\psi)\Delta g(r_P, Q)d\sigma_Q, \tag{5.61}$$

where $\Delta g(r_P, Q)$ is the gravity anomaly at point Q on the sphere of radius r_P, we can rewrite Eq. (5.60) as:

$$dN_{dwc} = \frac{T^* - \widetilde{T}}{\gamma_0} + \frac{\widetilde{T} - (T)}{\gamma_0} = \frac{R}{4\pi\gamma_0} \iint_{\sigma} S(\psi)\{[\Delta g^* - \Delta g(r_P, Q)] + [\Delta g(r_P, Q) - \Delta g]\} d\sigma = dN^B_{dwc} + dN^{te}_{dwc} \tag{5.62}$$

The first part we call "the spherical shell effect"; the second part we name "the terrain effect". dN^B_{dwc} can be evaluated by first decomposing it into two new parts:

$$dN^B_{dwc} = \frac{T^* - T}{\gamma_0} + \frac{T - \widetilde{T}}{\gamma_0} = dN^{B1}_{dwc} + dN^{B2}_{dwc}. \tag{5.63}$$

In order to simplify dN^{B1}_{dwc}, we make a Taylor expansion of T^* around the point P at the Earth's surface along the vertical through the location of T^*. The result is (now limited to the second power of H_P):

$$T^* \approx T_P - H_P \left(\frac{\partial T}{\partial H}\right)_P + \frac{H_P^2}{2} \left(\frac{\partial^2 T}{\partial H^2}\right)_P, \tag{5.64}$$

where H_P is the orthometric height. From the boundary condition of physical geodesy in spherical approximation, we obtain:

$$\left(\frac{\partial T}{\partial H}\right)_P = -\Delta g_P - 2\frac{T_P}{r_P}, \tag{5.65}$$

and, with the approximation $\partial/\partial r_P \approx \partial/\partial H_P$:

$$\frac{\partial^2 T}{\partial H^2}|_P \approx -\frac{\partial \Delta g}{\partial H}|_P + 2\frac{\Delta g_P}{r_P} + 6\frac{T_P}{r_P^2}. \tag{5.66}$$

Inserting Eqs. (5.65 and 5.66) into dN^{B1}_{dwc} of (5.22), one obtains:

$$dN^{B1}_{dwc} = \frac{T^* - T_P}{\gamma_0} \approx \frac{\Delta g_P H_P}{\gamma_0} + 2\frac{H_P}{r_P}\tilde{\zeta}_P - \frac{H_P^2}{2\gamma_0}\frac{\partial \Delta g}{\partial H}\bigg|_P + 3\left(\frac{H}{r}\right)^2_P \tilde{\zeta}_P - \frac{H_P^2 \Delta g_P}{r_P \gamma_0} \tag{5.67}$$

where we have introduced the symbol $\tilde{\zeta} = \gamma\zeta/\gamma_0$, with ζ being the height anomaly defined as the disturbing potential at the surface divided by normal gravity at normal height γ.

As the remaining part of Eq. (5.63) can be simplified to:

$$dN_{dwc}^{B2} = \frac{T_P - \widetilde{T}}{\gamma_0} = \left(1 - \frac{R}{r_P}\right)\tilde{\zeta}_P = \frac{H_P}{r_P}\tilde{\zeta}_P, \tag{5.68}$$

it follows that the Bouguer shell contribution to the dwc effect (limited to terms within second power of height, and omitting terms not contributing to the "1 cm geoid") becomes:

$$dN_{dwc}^{B} \approx \frac{H_P \Delta g_P}{\gamma_0} + 3\frac{H_P}{r_P}\zeta_P - \frac{H_P^2}{2\gamma_0}\left(\frac{\partial \Delta g}{\partial H_P}\right). \tag{5.69}$$

We now consider the last part of Eq. (5.62), "the terrain effect". As the argument includes a gravity-anomaly difference, which to first approximation can be written as:

$$\Delta g(r_P, Q) - \Delta g \approx (H_P - H)\left(\frac{\partial \Delta g}{\partial H}\right)_Q, \tag{5.70}$$

and Stokes' kernel is decreasing with the geocentric angle ψ, we may approximate the integration area to a cap around the computation point. The result is:

$$dN_{dwc}^{te} \approx \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S(\psi)(H_P - H)\left(\frac{\partial \Delta g}{\partial H}\right)_Q d\sigma_Q. \tag{5.71}$$

Equation (5.71) suffers from two approximations, namely that the Taylor series in derivatives of the surface gravity anomaly is limited to first degree, and that the integration area has been limited to a cap. The second problem is easily solved by extending the area of integration, while the first problem can be more difficult to solve.

In summary, the dwc effect for the original Stokes' formula becomes:

$$dN_{dwc} \approx dN_{dwc}^{B} + dN_{dwc}^{te}, \tag{5.72}$$

where the two components were given in Eqs. (5.69 and 5.71). The first two terms of dN_{dwc}^{B} are usually the most significant ones. For instance, the first term reaches 1 m, if the gravity anomaly is 200 mGal on top of a 5 km high mountain. For the height of Mt. Everest, it reaches 1.8 m. In practice, the height anomaly in Eq. (5.69) can be set to the geoid height determined from some approximate geoid model [The second term of Eq. (5.69) differs from Sjöberg (2003b; Eq. 12), which has a printing error. See also Ågren (2004b).].

The above correction holds also in the cases that the geoid is located (a) at the Earth's surface and (b) above the surface. In case a, the term dN_{dwc}^{B} vanishes so that

only dN^{te}_{dwc} remains. (In addition the topographic bias vanishes.) In case (b), one notices that H_P (the orthometric height at P) is negative, leading to the fact that dN^{B}_{dwc} corresponds to an upward continuation).

5.3.2 The Dwc Effect for the Modified Stokes' Formula

In the modified Stokes' formula, we may assume that Stokes' kernel is modified to some selected degree L, yielding the modified kernel function

$$S^L(\psi) = S(\psi) - \sum_{n=2}^{L} \frac{2n+1}{2} s_n P_n(\cos\psi), \tag{5.73}$$

where s_n are the modification parameters to be selected. Also, the integration area is usually limited, say, to a spherical cap σ_0 with geocentric angle ψ_0 around the computation point. As a result, the dwc effect becomes:

$$dN^{L}_{dwc} = \frac{R}{4\pi\gamma_0} \iint_{\sigma_0} S^L(\psi)[\Delta g^* - \Delta g] d\sigma. \tag{5.74}$$

Similarly to above, we may decompose (5.74) into a Bouguer shell effect and a terrain effect:

$$\begin{aligned} dN^{L}_{dwc} &= \frac{R}{4\pi\gamma_0} \iint_{\sigma_0} S^L(\psi)\{[\Delta g^* - \Delta g(r_P, Q)] + [\Delta g(r_P, Q) - \Delta g]\} d\sigma \\ &= dN^{B,L}_{dwc} + dN^{te,L}_{dwc} \end{aligned} \tag{5.75}$$

The Bouguer shell effect can be written as:

$$dN^{B,L}_{dwc} = dN^{B}_{dwc} + c \sum_{n=2}^{\infty} \left[\left(\frac{R}{r_P} \right)^{n+1} - 1 \right] \left(s^*_n + Q^L_n \right) \Delta g_n, \tag{5.76}$$

where dN^{B}_{dwc} is the same as in Eq. (5.63), and

$$s^*_n = \begin{cases} s_n & if \quad 2 \le n \le L \\ 0 & otherwise \end{cases}. \tag{5.77}$$

Q^L_n are the Molodensky truncation coefficients of the modified Stokes' formula (see, e.g. Sjöberg 2003a). In practice, Δg_n of the above sum is taken from an Earth gravitational model, which requires the upper limit of the sum be set equal to, or below, its maximum order.

The second term of (5.75) can be approximated by:

$$dN_{dwc}^{te,L} = \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S^L(\psi)[\Delta g(r_P, Q) - \Delta g]d\sigma$$
$$\approx \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S^L(\psi)(H_P - H_Q)\left(\frac{\partial \Delta g}{\partial H}\right)_Q d\sigma_Q \tag{5.78}$$

This completes the discussion on the dwc effect.

5.4 Atmospheric Corrections

As topographic masses, the atmospheric masses are "forbidden masses" that violate Stokes' formula, and their attraction on the gravity field must therefore be corrected for.

5.4.1 The IAG Approach

In the approach to atmospheric gravity reduction by the International Association of Geodesy (IAG), it is assumed that the Earth is spherical with a spherical layering of the atmosphere (Moritz 1980, p. 422). Hence, if M_a is the total mass of the atmosphere, and $M(r)$ and $m(r)$ are the masses outside and inside the sphere through the computation point at radius r, *the gravity correction,* or *direct atmospheric effect on gravity*, becomes:

$$\delta g^a = \frac{GM(r)}{r^2} = \frac{GM_a - Gm(r)}{r^2}. \tag{5.79}$$

By modelling the atmospheric density versus elevation, $M(r)$ can be computed and the direct effect applied to each gravity observation. The corresponding indirect effect on the geoid is small (about—7 mm) and therefore usually neglected.

Sjöberg (1999) and Sjöberg and Nahavandchi (2000) emphasized that the IAG approach leads to a very serious truncation bias if the method is not used with care. This is obvious from the following derivation.

The effect of Eq. (5.79), when downward continued to sea level, is:

$$\delta g^* = \gamma_a - \frac{Gm(r)}{R^2}, \tag{5.80}$$

where $\gamma_a \approx 0.86$ mGal is the atmospheric gravity at sea level. Inserting (5.80) into Stokes' formula, one obtains the direct atmospheric effect on the geoid height. As γ_a is a constant, it will not be harmful, as Stokes' (global) integral is blind to a constant gravity anomaly. However, this changes drastically if Stokes' formula is truncated,

Table 5.1 The atmospheric bias based on Eqs. (5.81a, 5.81b) for $\gamma_0 = 981$ Gal, $\gamma_a = 0.86$ mGal and $R = 6378$ km. Unit: metre

ψ_0°	0	3	10	25	39	50	75	100	125	150	180
δN_a	0	0.324	1.156	2.638	3.138	2.839	0.612	−1.667	−2.120	−0.994	0

say, to a cap σ_0. Disregarding the small contribution from the last term of (5.80), the direct effect on the geoid becomes:

$$\delta N^a_{dir} \approx -\frac{R}{4\pi}\frac{\gamma_a}{\gamma}\iint_{\sigma-\sigma_0} S(\psi)d\sigma = -\frac{R}{2}\frac{\gamma_a}{\gamma}Q_0(\psi_0), \tag{5.81a}$$

where:

$$Q_0(\psi_0) = -4t_1 + 5t_1^2 + 6t_1^3 - 7t_1^4 + 6t_1^2\left(1 - t_1^2\right)\ln\left(t_1 + t_1^2\right) \tag{5.81b}$$

is the zero-degree Molodensky truncation coefficient with $t_1 = \sin(\psi_0/2)$. As a result, Eq. (5.81a) yields a very significant bias, which depends on the integration cap size, having a maximum of 3.14 m for $\psi_0 = 39°$, which is much more significant than the effect to be corrected for. See Table 5.1.

5.4.2 The KTH Approach

Newton's volume integral for the atmospheric potential at point P can be written as:

$$V^a(P) = \iint_\sigma \int_{r_s}^{\infty} \frac{\rho^*(r)r^2 dr}{l_P} d\sigma \tag{5.82a}$$

where:

$$\rho^*(r) = G\rho_a(r) = \rho_0^*(R/r)^v. \tag{5.82b}$$

Here, ρ_0^* is atmospheric density at sea level times the gravitational constant, and $v > 2$ is a constant. As the atmosphere is located in the exterior of a surface point, the inverse distance can be extended by an internal type series

$$\frac{1}{l_P} = \frac{1}{r_P}\sum_{n=0}^{\infty}\left(\frac{r_P}{r}\right)^{n+1} P_n(t), \quad r_P < r. \tag{5.83}$$

After inserting Eqs. (5.82b and 5.83) into Eq. (5.82a), one obtains:

$$V^a(P) = \frac{\rho_0^* R^v}{r_P} \sum_{n=0}^{\infty} \iint_{\sigma} \int_{r_s}^{\infty} \left(\frac{r_P}{r}\right)^{n+1} r^{2-v} dr P_n(t) d\sigma, \tag{5.84}$$

or, after integration w.r.t. r and expanding $(R/r)^m \approx 1 - m(H/R)$ and considering the addition theorem of spherical harmonics, then:

$$V^a(P) \approx 4\pi\rho_0^* R^2 \left[\frac{1}{v-2} - \sum_{n=2}^{\infty} \frac{1}{2n+1} \left(\frac{r_P}{R}\right)^n \frac{H_n}{R}\right], \tag{5.85}$$

where

$$H_n = \frac{1}{4\pi} \iint_{\sigma} H Y_{nm} d\sigma. \tag{5.86}$$

Applying the boundary condition of physical geodesy, we obtain also the direct gravity anomaly effect:

$$\Delta g_{dir}^a(P) = \left(\frac{\partial}{\partial r_P} + \frac{2}{r_P}\right) V^a(P) \approx 4\pi\rho_0^* \left[\frac{2R}{v-2} - \sum_{n=2}^{\infty} \frac{n+2}{2n+1} H_n\right], \tag{5.87}$$

and, by inserting this into Stokes' formula with Stokes function in spectral form,

$$S(\psi) = \sum_{n=2}^{\infty} \frac{1}{n-1} \sum_{m=-n}^{\infty} Y_{nm}(P) Y_{nm}(Q), \tag{5.88}$$

and considering the orthogonality of the harmonics, the direct effect on the geoid height becomes:

$$N_{dir}^a(P) \approx \frac{4\pi\rho_0^* R}{\gamma} \sum_{n=2}^{\infty} \frac{n+2}{(2n+1)(n-1)} H_n. \tag{5.89}$$

Considering also that the indirect effect with $r_P = R$ and omitting the zero-degree effect, i.e.

$$\delta N_I^a \approx \frac{4\pi\rho_0^* R}{\gamma} \sum_{n=2}^{\infty} \frac{H_n}{2n+1}, \tag{5.90}$$

it follows that the total atmospheric effect becomes:

$$\delta N^a_{total} = \delta N^a_{dir} + \delta N^a_I \approx -\frac{4\pi\rho^*_0 R}{\gamma}\sum_{n=2}^{\infty}\frac{H_n}{n-1} = -\frac{\rho^*_0 R}{\gamma}\iint_{\sigma} S(\psi)Hd\sigma. \tag{5.91}$$

Here, the contribution from the zero- and first-degree terms (Sjöberg 2001), namely:

$$\left(\delta N^a_{total}\right)_{0,1} = -\frac{4\pi\rho^*_0 R^2}{\gamma(v-2)(v-3)}, \tag{5.92}$$

being of order −6 mm, is missing .

The atmospheric effect in the LSMSA is dependent on the type of EGM used in the modification. If the EGM is created only from terrestrial data, the atmospheric effects are the same as above. If the EGM includes satellite data to degree and order M and the modification is to degree $L > M$, then the total atmospheric effect becomes:

$$\delta N^{a;L}_{total} = \delta N^a_{dir} + \frac{4\pi\rho^*_0 R}{\gamma}\sum_{n=2}^{M}\left(Q^L_n + s_n\right)\frac{n-1}{2n+1}H_n, \tag{5.93}$$

where δN^a_{total} was given in Eq. (5.91).

5.5 Ellipsoidal Corrections

The well-known Stokes' formula determines the geoidal height from surface gravity anomalies on a sphere. As the Earth's shape is rather ellipsoidal with a flattening of the order of 1/300, it can be expected that the error in Stokes' formula also amounts to this order, corresponding to an error within several decimetres of the geoidal height.

Through the years, many authors have studied the ellipsoidal correction to Stokes' formula [see, e.g. the reference list of Sjöberg (2003c)], and each investigator seems to come up with her/his own, more or less unique, solution. For example, Huang et al. (2003) carried out a numerical comparison between the solutions of Molodensky et al. (1962), Moritz (1980), Martinec and Grafarend (1997) and Fei and Sideris (2000), and they found that all compared solutions disagree. Another solution to the ellipsoidal effect to Stokes' formula was presented by Martinec (1998a, Chap. 11, 1998b), but this solution is incomplete, as it solves the problem for the sphere with the gravity anomaly given on a sphere instead of the geodetic problem with gravity anomaly given on the Earth ellipsoid and the potential requested on the ellipsoid.

Sjöberg (2003c) used Green's formula to derive an integral solution of the ellipsoidal correction to Stokes' formula. The solution was also developed into a series of spherical harmonics. See also Sjöberg (2004b).

Nevertheless, today, the determination of the geoid from gravimetric data is most frequently performed as a combination of a global geopotential model (EGM) and surface gravity anomalies in a limited area around the computation point as a modified version of Stokes' formula. As the EGM can be correctly applied at sea level (approximated by the Earth ellipsoid), the remaining ellipsoidal correction to Stokes' formula concerns only the limited area of integration. If this area is sufficiently small, one would intuitively assume that the ellipsoidal correction be negligible [see Ellmann and Sjöberg (2004)].

5.5.1 *Components of the Ellipsoidal Correction of Stokes' Formula*

Following Sjöberg (2003d), we start to present the different components of the ellipsoidal correction to the geoid height, and in Sect. 5.5.2 we derive them as a Stokes'-type of integral, as well as a series of spherical harmonics.

Stokes' formula is approximately given by:

$$N^0 = \frac{R}{4\pi\gamma}\iint_{\sigma} S(\psi)\Delta g d\sigma, \tag{5.94}$$

where the gravity anomaly (Δg), here assumed to be downward continued to the sphere of radius R, is only approximately consistent with the spherical approximation of the "boundary condition", whose solution is Stokes' integral, and it must therefore be corrected to such a gravity anomaly (Δg^0) by the formula (limited to order e^2; Jekeli 1981; Cruz 1986)

$$\Delta \mathrm{g}^0 = \Delta \mathrm{g} - \mathrm{e}^2 \sin\theta\, \cos\theta \frac{\partial \mathrm{T}}{\mathrm{a}\partial\theta} - \mathrm{e}^2\left(3\cos^2\theta - 2\right). \tag{5.95}$$

We will assume that Δg (and therefore also Δg^0) are reduced to sea level, i.e. to the reference ellipsoid of radius $r_e = a\sqrt{1 - e^2\cos^2\theta}$, while for the Stokes' integration, performed on the sphere of radius a, we need to upward continue Δg^0 by the approximate gravity anomaly

$$\Delta G^0 = \Delta g^0 + (a - r_e)\left(\frac{\partial \Delta g^0}{\partial r}\right)_{r=a} = \Delta g + \delta G, \tag{5.96a}$$

where:

$$\begin{aligned}\delta G &= \Delta G^0 - \Delta g \\ &= e^2\left[\frac{a}{2}\cos^2\theta\left(\frac{\partial \Delta g}{\partial r}\right)_{r=a} - \sin\theta\cos\theta\left(\frac{\partial T}{a\partial\theta}\right)_{r=a} + (2-3\cos^2\theta)\frac{T_a}{a}\right] \end{aligned} \quad (5.96b)$$

is the needed correction to the original gravity anomaly to be consistent with Stokes' formula (on the sphere of radius a).

Finally, after Stokes' integration, the disturbing potential needs to be downward continued to the geoid, which corresponds to the last component in the following formula for the total ellipsoidal correction to Stokes' formula:

$$\delta N^e_{total} = kN^0 + \frac{a}{4\pi\gamma}\iint_\sigma S(\psi)\delta G d\sigma - e^2 a\frac{\cos^2\theta}{2\gamma}\left(\frac{\partial T}{\partial r}\right)_{r=a}, \quad (5.97)$$

where $k = (a-R)/R$ is a scale factor in augmenting N^0 to the sphere of radius R.

To summarize, the ellipsoidal correction to Stokes' formula consists of three components:

- The original gravity anomaly Δg needs both a geometrical and physical correction to become Δg^0, which is consistent with the boundary condition.
- Carrying out the Stokes' integration on the sphere of radius a rather than R, which yields the corrections kN^0 to Stokes' formula and $(a - r_e)\partial\Delta g^0/\partial r$ to δG.
- The conversion of the disturbing potential between the spheres of radii a and r_e causes the last term of (5.97).

5.5.2 *The Ellipsoidal Correction as a Harmonic Series and a Stokes' Integral*

Next we derive the ellipsoidal correction both as a harmonic series and as a Stokes' integral, convenient for applications.

Let us start with the following harmonic series expansion of the disturbing potential:

$$T = \frac{GM}{a}\sum_{n=2}^{\infty}\left(\frac{a}{r}\right)^{n+1}\sum_{m=-n}^{n} C_{nm}Y_{nm}(\theta,\lambda). \quad (5.98)$$

Substituting Eq. (5.98) into the gravity anomaly correction δG of Eq. (5.96b) and considering the following relations (Moritz 1980; Martinec 1998a):

$$\cos^2\theta Y_{nm} = E_{nm}Y_{n+2,m} + F_{nm}Y_{nm} + G_{nm}Y_{n-2,m} \tag{5.99a}$$

and

$$\sin\theta\,\cos\theta\frac{\partial}{\partial\theta}Y_{nm} = A_{nm}Y_{n+2,m} + B_{nm}Y_{nm} + D_{nm}Y_{n-2,m}, \tag{5.99b}$$

where (Sjöberg 2003c, 2004b):

$$\begin{aligned} A_{nm} &= nE_{nm},\ B_{nm} = -\frac{n(n+1)-3m^2}{(2n-1)(2n+3)} \\ D_{nm} &= -\frac{n+1}{2n-1}\sqrt{\frac{\left[(n-1)^2-m^2\right](n^2-m^2)}{(2n-3)(2n+1)}} \\ E_{nm} &= \frac{1}{2n+3}\sqrt{\frac{\left[(n+1)^2-m^2\right]\left[(n+2)^2-m^2\right]}{(2n+1)(2n+5)}} \\ F_{nm} &= -\frac{2}{3}B_{nm} + \frac{1}{3}, \end{aligned} \tag{5.99c}$$

one obtains the Laplace series for the correction δG:

$$\delta G = \sum_{n=2}^{\infty}\delta G_n, \tag{5.100a}$$

where:

$$\begin{aligned} \delta G_n = e^2\frac{GM}{2a}\sum_{m=-n}^{n} Y_{nm}\big[&\{3-(n^2+n+1)F_{nm}\}C_{nm} \\ &-(n^2-n-2)G_{nm}C_{n-2,m} - (n^2+3n+4)E_{nm}C_{n+2,m}\big] \end{aligned} \tag{5.100b}$$

Moreover, taking the radial derivative of the series of T of Eq. (5.98) and considering Eq. (5.99a), the last term of Eq. (5.97) can be expressed as:

$$\begin{aligned} \delta N^d = \frac{1}{\gamma}\sum_{n=2}^{\infty}\delta T_n^d\frac{e^2GM}{2a\gamma}\sum_{n=2}^{\infty}\frac{1}{n-1}\sum_{m=-n}^{n} Y_{nm}\big[&(n+1)F_{nm}C_{nm} \\ &+(n-1)G_{nm}C_{n-2,m} + (n+3)E_{nm}C_{n+2,m}\big] \end{aligned} \tag{5.101}$$

By summing up, one arrives at the following spectral solution for Eq. (5.97):

$$\delta N^{e}_{total} = kN^{0} + \frac{1}{\gamma}\sum_{n=2}^{\infty}\left(\frac{a\delta G_n}{n-1} + \delta T^{d}_{n}\right) = kN^{0} + \frac{a}{\gamma}\sum_{n=2}^{\infty}\frac{\delta g^{e}_{n}}{n-1}, \tag{5.102a}$$

where:

$$\delta g^{e}_{n} = \delta G_n + (n-1)\frac{\delta T^{d}_{n}}{a} \tag{5.102b}$$

Here, δg^{e}_{n} are the Laplace harmonics of the ellipsoidal correction to the gravity anomaly, which can be explicitly decomposed into the series (Sjöberg 2003c, 2004b):

$$\delta g^{e}(\theta,\lambda) = \sum_{n=2}^{\infty}\delta g^{e}_{n}(\theta,\lambda) = \sum_{n=2}^{\infty}\sum_{m=-n}^{n}\delta g^{e}_{nm}Y_{nm}(\theta,\lambda), \tag{5.103a}$$

with the coefficients:

$$\delta g^{e}_{nm} = e^{2}\frac{GM}{2a^{2}}\left\{[3-(n+2)F_{nm}]C_{nm} - (n+1)G_{nm}C_{n-2,m} - (n+7)E_{nm}C_{n+2,m}\right\}. \tag{5.103b}$$

As an alternative, Eq. (5.102a) can be written as a Stokes' formula in the space domain:

$$\delta N^{e}_{total} = \frac{R}{4\pi\gamma}\iint_{\sigma} S(\psi)\left(k\Delta g + \frac{a}{R}\delta g^{e}\right)d\sigma. \tag{5.104}$$

In case of the modified Stokes formula (e.g. LSMSA), we may assume that the contribution with a harmonic series is directly targeted at radius $r_p = r_e$, so that the total ellipsoidal correction follows in accord with (5.106) as:

$$\delta N^{e,L}_{total} = \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^{L}(\psi)\left(k\Delta g + \frac{a}{R}\delta g^{e}\right)d\sigma, \tag{5.105}$$

which can also be converted to the harmonic series

$$\delta N^{e,L}_{total} = \frac{R}{2\gamma}\sum_{n=2}^{\infty}\left(\frac{2}{n-1} - S^{*}_{n} - Q^{L}_{n}\right)(k\Delta g_n + \frac{a}{R}\delta g^{e}_{n}). \tag{5.106}$$

If the cap size is chosen as a few degrees only, the ellipsoidal correction to Stokes' modified formula is usually within a few millimetres.

5.6 Corrections in Quasigeoid Determination

As stated in the introduction to this chapter, the above corrections mainly concern geoid determination. In particular, there is no topographic correction in quasigeoid determination. (However, a dwc correction is still needed.) Nevertheless, sometimes (e.g. in the RCR technique), a direct topographic effect (DITE) is introduced to smooth the gravity anomaly under Stokes' integral, and, importantly, a corresponding PITE must then be added to the final result.

The atmospheric and ellipsoidal corrections in quasigeoid determination are practically the same as for geoid determination.

If the quasigeoid is estimated by the extended Stokes' formula, Eq. (7.13a) in a RCR approach, the correction for downward continuation of the gravity anomaly to the sphere of computation selected at sea level (e.g. by solving Poisson's integral equation) is the same as for geoid determination. Ågren et al. (2009) applied the dwc correction for ζ as a Stokes' integral on the sphere at sea level radius in the LSMSA approach. On the other hand, if Stokes' integration is applied on a sphere at point level, the correction for analytical continuation (upward and/or downward) refers to the sphere through the computation point.

The additive corrections in the LSMSA method are shortly given in Sect. 6.3.2.

References

Ågren J (2004a) The analytical continuation bias in the geoid determination using potential coefficients and terrestrial gravity data. J Geod 78:314–332

Ågren J (2004b) Regional geoid determination methods for the era of satellite gravimetry, numerical investigations using synthetic Earth gravity models. Doctoral thesis in Geodesy, Royal Institute of Technology, Stockholm, Sweden

Ågren J, Sjöberg LE, Kiamehr R (2009) The new gravimetric quasigeoid model KTH08 over Sweden. J Appl Geod 3(3):143–153

Cruz JY (1986) Ellipsoidal corrections to potential coefficients obtained from gravity anomaly data on the ellipsoid. Rep 371, Department of Geodetic Science and Surveying, Ohio State University, Columbus

Ellmann A, Sjöberg LE (2004) Ellipsoidal correction for the modified Stokes formula. Boll Geod Sci Aff 63:153–172

Fei ZL, Sideris MG (2000) A new method for computing the ellipsoidal correction for Stokes's formula. J Geod 74:223–231

Heiskanen WA, Moritz H (1967) Physical Geodesy. W H Freeman and Co., San Francisco and London

Huang J, Véronneau M, Pagiatakis SD (2003) On the ellipsoidal correction to the spherical Stokes solution of the gravimetric geoid. J Geodesy 77(3–4):171–181

Jekeli C (1981) The downward continuation to the earth's surface of truncated spherical and ellipsoidal harmonic series of the gravity height anomaly. Rep 323, Department of Geodetic Science and Surveying, OSU, OH

Martinec Z (1998a) Boundary-value problems for gravimetric determination of a precise geoid. Lecture notes in earth sciences. Springer, Berlin

Martinec Z (1998b) Construction of Green's function for the Stokes boundary-value problem with ellipsoidal corrections in the boundary condition. J Geodesy 72:460–472
Martinec Z, Grafarend EW (1997) Solution to the Stokes boundary-value problem on an ellipsoid of revolution. Stud Geophys Geod 41:103–129
Molodenskii MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitation field and figure of the earth. Translated from Russian (1960), Israel Program for Scientific Translations, Jerusalem
Moritz H (1980) Advanced physical geodesy. Herbert Wichmann Verlag, Karlsruhe
Moritz H (2003) The strange behavior of asymptotic series in mathematics, celestial mechanics and physical geodesy. In: Grafarend EW, Krumm FW, Schwarze VS (eds) Geodesy: the challenge of the third millennium. Springer, Berlin, pp 371–377 (in the Internet: Google-Suchenach "asymptotic series")
Sjöberg LE (1977) On the errors of spherical harmonic developments of gravity at the surface of the earth, Department of Geodetic Science Rep No. 257, The OSU, Columbus, OH
Sjöberg LE (1999) The IAG approach to the atmospheric geoid correction in Stokes formula and a new strategy. J Geodesy 73(7):362–366
Sjöberg LE (2000) On the topographic effects by the Stokes-Helmert method of geoid and quasigeoid determinations. J Geod 74(2):255–268
Sjöberg LE (2001) Topographic and atmospheric corrections of the gravimetric geoid determination with special emphasis of the effects of degrees zero and one. J Geod 75:283–290
Sjöberg LE (2003a) A computational scheme to model the geoid by the modified Stokes's formula without gravity reductions. J Geod 77(2003):423–432
Sjöberg LE (2003b) A solution to the downward continuation effect on the geoid determined by Stokes' formula. J. Geod 77:94–100
Sjöberg LE (2003c) The ellipsoidal corrections to order e^2 of geopotential coefficients and Stokes' formula. J Geod 77:139–147
Sjöberg LE (2003d) The correction to the modified Stokes formula for an ellipsoidal Earth. In Santos M (ed) Honoring the academic life of Petr Vanicek, UNB Technical Report No. 218, New Brunswick, pp 99–110
Sjöberg LE (2004a) The effect on the geoid of lateral topographic density variations. J Geod 78:34–39
Sjöberg LE (2004b) A spherical harmonic representation of the ellipsoidal correction to the modified Stokes formula. J Geod 78(2004):180–186
Sjöberg LE (2005) Discussion on the approximations made in the practical implementation of the remove-compute-restore technique in regional geoid modelling. J Geod 78:645–653
Sjöberg LE (2007) The topographic bias by analytical continuation in physical geodesy. J Geod 81:345–350
Sjöberg LE (2008) Answers to the comments by M. Vermeer on L.E. Sjöberg (2007) "The topographic bias in physical geodesy". J Geod 81:345–350. doi:10.1007/s00190-006-0112-2
Sjöberg LE (2009a) The terrain correction in gravimetric geoid determination—is it needed? Geophys J Int 176:14–18
Sjöberg LE (2009b) On the topographic bias in geoid determination by the external gravity field. J Geod 83:967–972
Sjöberg LE (2009c) Solving the topographic bias as an initial value problem. Art Sat 44(3):77–84
Sjöberg LE (2014) On the topographic effects by Stokes' formula. J Geod Sci 4:130–135
Sjöberg LE (2015a) The secondary indirect topographic effect in physical geodesy. Stud Geophys Geod 59:173–187
Sjöberg LE (2015b) The topographic bias in Stokes' formula vs. the error of analytical continuation by an earth gravitational model—are they the same? J Geod Sci 5:171–179
Sjöberg LE, Nahavandchi H (2000) The atmospheric geoid effects in Stokes's formula. Geophys J Int 140:95–100
Vermeer M (2008) Comments on L. E. Sjöberg (2007) "The topographic bias by analytical continuation in physical geodesy". J Geod 81:345–350

Chapter 6
Applications and Comparisons of LSMSA and RCR

Abstract In this chapter, geoid determination by the remove-compute-restore (RCR) technique and Least Squares Modification of Stokes' formula with Additive corrections (LSMSA) are briefly presented. The basic formulas of each method are developed, followed by a theoretical comparison. The advantages of the LSMSA method include: (a) a unique spectral least squares matching of errors of gravity and EGM data, as well as truncation of the integration area to a cap; (b) additive corrections that are easier to compute than direct and indirect effects; (c) the downward continuation effect is more stable; (d) the bias in the atmospheric effect in the standard IAG formula is avoided; and (e) each additive correction is easily updated whenever new data is available. The chapter ends with some case studies with numerical results. In most cases, even in international comparisons with other techniques, the LSMSA method provides the best agreement with geoid determination by the independent GPS-levelling technique.

Keywords Additive corrections · Least squares modification · Remove-Compute-Restore

6.1 Introduction

Below we only describe the determination of the geoid and quasigeoid by the methods of remove-compute-restore (RCR) and Least Squares Modification of Stokes' formula with Additive corrections (LSMSA), which we believe are the most successful techniques for geoid determination to date. We compare these methods, both theoretically and practically in various applications taken from the literature. In all formulas below it is assumed that L (the degree of modification) is larger than or equal to M (the selected upper limit of degree of the EGM). If $L < M$, the formulas hold with some minor changes. (In this way the method can be used with ultra-high degree EGMs. See Sect. 4.4.6.)

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_6

6.2 Geoid Determination

As discussed in Chap. 5, geoid determination by Stokes' formula requires that: (a) the effects of all masses above sea level (approximated by a sphere of radius R), i.e. the masses of the topography and atmosphere, be removed [*direct topographic and atmospheric effects on gravity anomaly*; $\delta\Delta g^T_{dir}(=\text{DITE on }\Delta g)$ and $\delta\Delta g^a_{dir}$]; (b) the surface gravity anomaly be downward continued to sea level (*the dwc effect on gravity anomaly* $\delta\Delta g_{dwc}$); and (c) *the ellipsoidal effect on the gravity anomaly* be accounted for by $\delta\Delta g^e_{dir}$.

After applying Stokes' formula, the effects of topography, atmosphere and ellipsoidal must be restored, yielding *the indirect topographic, atmospheric and ellipsoidal effects on the geoid height*, $\delta N^T_I = \text{PITE}$, δN^a_I and δN^e_I. All these corrections were described in detail in Chap. 5. However, there are some practical differences in applying the corrections in the RCR method and in the LSMSA method that will be described in the following subsections.

6.2.1 Remove-Compute-Restore Technique

In the RCR method, the topographic, atmospheric and ellipsoidal corrections are applied to the surface-gravity anomaly, and the corrected anomaly is downward continued to sea level (approximated by the sphere of mean sea-level radius R) to make it ready for the modified Stokes' integration by the estimator:

$$\begin{aligned}\tilde{N}^{L,M} = &\frac{(T_0)_g + (T_1)_g + (\delta T_0 + \delta T_1)_{dir}}{\gamma_0} \\ &+ \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S^L(\psi)\left[\Delta g + \delta\Delta g^T_{dir} + \delta\Delta g^a_{dir} + \delta\Delta g^e_{dir}\right]^* d\sigma \\ &+ c\sum_{n=2}^{M}\left(Q^L_n + s^*_n\right)\left[\Delta g^{EGM}_n + \left(\delta\Delta g^T_n\right)_{dir}\right] + \delta N^T_I + \delta N^a_I + \delta N^e_I, \end{aligned} \tag{6.1}$$

where all symbols were defined above and in Chaps. 4 and 5. Note that the estimator is here augmented by the zero- and first-degree disturbing potential harmonics to provide the absolute geoid height. The direct effects $\left(\delta\Delta g^T_n\right)_{dir}$ are needed to enable applying the EGM below the topography. The analytical continuation (e.g. by solving Poisson's integral) brings the bracket []* down to the mean Earth sphere. Many of the methods in the literature for geoid determination based on the modification of Stokes' formula use Eq. (6.1) in one way or another, and they mainly differ in the way they modify Stokes' formula, i.e. in the way of selecting integration area and modification parameters s_n, as well in formulas chosen for the direct and indirect effects. They also differ in being more or less rigorous in

applying the direct and indirect corrections. All these methods can be called Remove-Compute-Restore (RCR) techniques: All forbidden effects on the gravity anomaly should be removed prior to Stokes' integration, and they should be restored on the geoid height after integration.

Frequently one goes one step further along the remove-restore line by using *a residual gravity anomaly by a higher-order reference field.* (For a review of the UNB technique, see Ellmann and Vanicek 2007). Then, starting from Eq. (4.27a), one obtains the estimator

$$\begin{aligned}\tilde{N}^{L,M} = \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S^L(\psi)\left[\Delta g^M + \delta\Delta g^{T,M}_{dir} + \delta\Delta g^{a}_{dir} + \delta\Delta g^{e}_{dir}\right]^* d\sigma \\ + \sum_{n=0}^{M} N_n^{EGM} + \delta N^{T,M}_{dir} + \delta N^{T}_{I} + \delta N^{a}_{I} + \delta N^{e}_{I},\end{aligned} \tag{6.2a}$$

where the direct effects under the integral are given as corrections to the residual surface gravity anomaly. For instance, in this case the direct topographic effect becomes:

$$\delta\Delta g^{T,M}_{dir} = -\mu R \sum_{n=M+1}^{\infty}\left(\frac{R}{r_Q}\right)^{n+2}(n-1)a_n(Q) \tag{6.2b}$$

with a_n given by Eq. (5.2b), and Q is on the surface at the integration point. The direct topographic effect for the sum of harmonics N_n^{EGM} is given by $\delta N^{T,M}_{dir}$.

The main advantage of the RCR method is that Stokes' formula operates with a residual gravity anomaly (reduced both for topographic effects and for the long-wavelength gravity field). The DITE and PITE can be further reduced by a topographic compensation model (e.g. by an isostatic model). Note that, similar to the LSMSA method, this method is also suitable for a least squares choice of the modification parameters. However, this is seldom or never performed in practice, but the modification parameters, if used at all, are mostly taken from some deterministic technique such as Wong and Gore (1969).

6.2.2 Least Squares Modification of Stokes' Formula with Additive Corrections (LSMSA)

The *Least Squares Modification of Stokes' Formula with Additive corrections (LSMSA)*, also called the KTH method, means that one of the least squares estimators of Sect. 4.4.4 is used for selecting the modification parameters s_n by considering the errors of gravity data, the EGM and the truncation of Stokes' integral, and the gravity anomaly is not corrected for the direct effects as in Eq. (6.1), but the direct effects on the geoid height are added to the indirect effects, yielding *the combined or total corrections/effects.* Hence, Eq. (6.1) becomes:

$$\tilde{N}^{L,M} = \frac{(T_0 + T_1)_g}{\gamma_0} + \frac{R}{4\pi\gamma_0}\iint_{\sigma_0} S^L(\psi)\Delta g d\sigma + c\sum_{n=0}^{M}\left(Q_n^L + s_n\right)\Delta g_n^{EGM} + \delta N_{comb}^T + \delta N_{dwc}^L + \delta N_{tot}^a + \delta N_{tot}^e, \tag{6.3a}$$

where $\delta N_{comb}^T = \delta N_{corr}^T$ is the combined topographic effect as given approximately by Eq. (5.16) and precisely by (5.51). δN_{tot}^a and δN_{tot}^e are the total atmospheric and ellipsoidal effects given by Eqs. (5.93) and (5.104) or (5.105), respectively. Finally, the dwc correction is given as in Sect. 5.3.2 by:

$$dN_{dwc}^L \approx \frac{H_p\Delta g_p}{\gamma_0} + 3\frac{H_p}{r_p}\zeta_p - \frac{H_p^2}{2\gamma_0}\left(\frac{\partial\Delta g}{\partial H_p}\right) + c\sum_{n=2}^{\infty}\left(Q_n^L + s_n^*\right)\left[\left(\frac{R}{r_p}\right)^{n+1} - 1\right]\Delta g_n + c\iint_{\sigma_0} S^L(\psi)\left(H_p - H_Q\right)\left(\frac{\partial\Delta g}{\partial h}\right)_Q d\sigma_Q. \tag{6.3b}$$

Hence, the first row of Eq. (6.3a) is the modified Stokes' formula, which uses the original EGM and gravity anomaly data. The second row of the equation consists of the additive corrections. These combined and total corrections were derived in Chap. 5.

6.3 Quasigeoid Determination

The determination of the quasigeoid height is very similar to that of determining the geoid height, except that the topographic corrections DITE and PITE in the remove-restore technique or combined topographic effect in the LSMSA method are not necessary, but still used in one way or another in the RCR technique to smooth the residual gravity anomaly. Also, Stokes' original function is often replaced by the extended Stokes' function, but not necessarily (see below).

6.3.1 The RCR Technique

Similar to Eq. (6.2a), which uses the original Stokes function for geoid determination, the quasigeoid can be determined by the extended Stokes' function:

$$\zeta^{L,M} = \frac{(T_0)_p + (\delta T_1)_p}{\gamma} + \frac{R}{4\pi\gamma}\iint_{\sigma_0} S^L(r_p,\psi)\left[\Delta g^M + \delta\Delta g^T_{dir} + \delta\Delta g^a_{dir} + \delta\Delta g^e_{dir}\right]^* d\sigma$$
$$+ \sum_{n=0}^{M} \zeta_n^{EGM} + \delta\zeta^{T,M}_{dir} + \delta\zeta^T_I + \delta\zeta^a_I + \delta\zeta^e_I, \tag{6.4}$$

where $\tilde{c} = R/(2\gamma) = c\gamma_0/\gamma$,, and the DITE $\left(\delta\Delta g^T_{dir}\right)$ and PITE $(\delta\zeta^T_I)$ are not needed but (as mentioned previously) is usually applied in one way or another to reduce the argument under Stokes' integral. If the gravity anomaly (both under the integral and in the EGM) is smoothed with some topographic model, it is important that a consistent PITE is introduced to avoid introducing a bias in the result. The downward continued residual gravity anomaly $[\,]^*$ implies that the atmospheric and ellipsoidal effects are the same as for geoid height determination (except that Bruns' formula is now applied with normal gravity γ at normal height instead of γ_0 at the reference ellipsoid).

Conveniently, the Stokes' integration can also be applied at point level (see Sect. 7.1.3), yielding the formula

$$\bar{\zeta}^{L,\mathrm{M}} = (\zeta^{L,\mathrm{M}}) + \sum_{n=0}^{M} \zeta_n^{EGM} + \delta\zeta^{T,M}_{dir} + \delta\zeta^{L,\mathrm{M}}_{dwc} + \delta\zeta^T_I + \delta\zeta^a_I + \delta\zeta^e_I, \tag{6.5a}$$

where:

$$\left(\zeta^{L,\mathrm{M}}\right) = \frac{r_P}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)[\,]\,d\sigma + \sum_{n=0}^{M} \zeta_n^{EGM} \tag{6.5b}$$

and

$$\left(\delta\zeta^{L,\mathrm{M}}_{dwc}\right) \approx \frac{r_P}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)(H_P - H_Q)\left(\frac{\partial[\,]}{\partial h}\right)_Q d\sigma_Q, \tag{6.5c}$$

where:

$$[\,] = \Delta g^M + \delta\Delta g^T_{dir} + \delta\Delta g^a_{dir} + \delta\Delta g^e_{dir}. \tag{6.5d}$$

6.3.2 The LSMSA Method

In the LSMSA technique, the height anomaly can be expressed as:

$$\tilde{\zeta}^{L,M} = \frac{(T_p)_0^{EGM} + (T_p)_1^{EGM}}{\gamma} + \frac{r_p}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)\Delta g d\sigma + \frac{r_p}{2\gamma}\sum_{n=2}^{M}\left(Q_n^L + s_n\right)\left(\frac{R}{r_p}\right)^{n+1}\Delta g_n^{EGM} + \delta\zeta_{dwc}^{L,M} + \delta\zeta_{tot}^{a} + \delta\zeta_{tot}^{e}, \tag{6.6a}$$

where $\delta\zeta_{tot}^{a}$ and $\delta\zeta_{tot}^{e}$ are the same as the total atmospheric and ellipsoidal corrections for the geoid height (except that they are now applied at point level and not at geoid level, and the scaling is for γ instead of γ_0), and

$$\delta\zeta_{dwc}^{L,M} \approx \frac{r_p}{4\pi\gamma}\iint_{\sigma_0} S^L(\psi)(H_P - H_Q)\left(\frac{\partial\Delta g}{\partial h}\right)_Q d\sigma_Q. \tag{6.6b}$$

6.4 A Theoretical Comparison of the RCR and LSMSA Methods

Both the RCR and LSMSA techniques can be applied more or less rigorously with all direct and indirect effects included, and the same modification parameters can be chosen (e.g. according to least squares based on stochastic observational errors with known error-degree variances). Theoretically, both approaches should lead to the same results for identical selection of modification parameters. In practice, this is not likely to be the case for a number of reasons to be specified subsequently.

Some advantages of the LSMSA method versus the RCR strategy:

- LSMSA has the advantage to the RCR method that it uses least squares to minimize the effects of errors in the data and truncation. (This advantage goes away if RCR is based on the same principle.)
- Most additive corrections used in LSMSA are easier, some much easier, to compute than the corresponding direct and indirect effects in the RCR method. Examples include:
  - As direct and indirect effects are added directly on the geoid height, there are computational advantages.
  - The topographic bias used in the LSMSA method is most simple to compute. On the contrary, the rigorous determination of direct topographic effect on the gravity anomaly in the RCR technique becomes unstable for high topography and high-resolution demands on the solution.
  - The LSMSA technique needs no terrain correction in contrast to RCR techniques (Sjöberg 2009).
- LSMSA avoids the global integration needed in the strict formula in determining the DITE in the RCR technique.
- The dwc effect is much more stable in the LSMSA technique (Sect. 5.3).

- The bias in the atmospheric effect of RCR is avoided in the LSMSA method.
- The LSMSA method is more flexible, as the additive corrections can be added/changed whenever the data for them becomes available/improved. (It means that the repetition of the main computational steps can be avoided.)
- If the RCR method is not applied with the same rigour in the direct and indirect effects, a bias will occur. This problem is avoided in the LSMSA method.
- RCR is frequently applied without a modification of Stokes' kernel function, which requires a large area of integration not to introduce a truncation error.

A disadvantage of LSMSA:

- As the LSMSA technique works on the full gravity anomaly (without any reduction), the interpolation and integration of the anomaly is more challenging than in the RCR method applied with a residual gravity anomaly. This problem needs special attention. (For example, a simple solution is to use RCR technique for the interpolation, but in this case no special care is needed to make global integrations, which is practically advantageous.)

For more details, see, e.g. Sjöberg (2003).

Further considerations:

- Note the correction term $\delta N^{T}_{dir,M}$, a most significant term, which is usually omitted in applications of the RCR technique. It is the topographic correction of the EGM.
- Until now (2016), the IAG Geoid School only practiced this technique without (i.e. with $L = 0$), or with simple modification of Stokes' kernel, and some correction terms are simplified or omitted.
- In practicing the RCR method (in contrast to LSMSA), there is usually no specific limitation of the area of integration, but all available gravity data are used in Stokes' integral. For a large integration area, the solution will suffer from the systematic errors in the gravity anomaly data (which problem may be reduced by an appropriate modification of Stokes' function).
- The modification by Vanicek-Kleusberg (1987) uses $L = M$ and Molodensky type of modification parameters (i.e. the minimum error limit of truncation error).

One may say that the LSMSA method uses the rigorous RCR-technique from Eqs. (6.1) and (6.4), but direct and indirect effects are added to become the combined or total effects.

Summary of the comparison:

- The LSMSA uses optimal (least squares) matching of various error sources. No such criterion is standard in applications of the RCR-technique.
- The LSMSA is numerically efficient. (Note: RCR requires both direct and indirect effects.) In particular, the combined topographic effect is simple in LSMSA (see Sect. 5.1).

- The dwc effect on the geoid height (used in LSMSA) is numerically stable (Sect. 5.2), while the dwc effect on gravity anomaly (used in RCR technique) is much more ill-conditioned.
- There is no bias in the atmospheric effect (LSMSA; Sect. 5.3). This may be a problem in the RCR technique.
- The RCR technique works with residual gravity anomalies, while LSMSA uses the full gravity anomaly. This calls for special technique in the LSMSA to achieve this advantage of the RCR method.
- The RCR technique should not be used without modifying the kernel function. In this respect, most important are the demands on filtering out the long-wavelength errors from the gravity data and reducing the truncation error.

6.5 Practical Experiences of LSMSA

Here we discuss some of the practical experiences gained in using the KTH method.

6.5.1 The Choice of Error Degree Variances

The error-degree variances of the data as well as the signal degree variances are used to tune the weighting of the data in order to minimize the influence of data errors, as well as the inevitable error caused by truncation of integration area and/or the series of the EGM. This (least squares) weighting takes place in the spectrum of the data, and it is ruled by the error and signal power spectra. As with standard least squares adjustment, the estimated unknowns (in our case the geoid heights) are only secondarily influenced by the choice of errors, while the estimated accuracies of unknowns are directly affected by this choice.

The error-degree variances of the EGM are frequently provided directly along with the EGM coefficients. However, there qualities may be doubtful, in particular at higher degrees. The major problem is, nevertheless, the error spectrum of the gravity anomaly data. This spectrum is not directly available, but must be estimated in one way or another. The LSMSA method assumes that all errors are random with expectation zero, but gravity data usually include systematic errors, which show up as correlations of the data. Ideally, such systematic effects should be removed prior to adjustment, and much care should be spent on cleaning the data from gross and systematic errors. However, remaining data will still include such undetected errors.

In practice, one usually assumes that the covariance function of the data, $C(\psi)$ is (globally) homogeneous and isotropic, and using the simple representation

$$C(\psi) = \sum_{n=2}^{\infty} \sigma_n^2 P_n(\cos\psi), \tag{6.7a}$$

it implies that, once the degree variances σ_n^2 have been fixed, it is merely a function of geocentric angle between the data points. To assume that this global covariance function is representative for the true covariance function is, of course, questionable, but, from a practical point of view, we only have to regard it as a model with parameters chosen for a specific geographic area of application. The most important parameters of this covariance function are the variance $C(0)$ and its correlation length ψ_c. The latter parameter is defined as that geocentric angle where the covariance has decreased to half of the magnitude of the variance, i.e.

$$C(\psi_c) = C(0)/2. \tag{6.7b}$$

Having got this information we usually can model the covariance function and the degree variances.

Sjöberg (1986) gave a simple example, which we present here. By *assuming* that the covariance function is of the closed form ("the reciprocal distance type")

$$C(\psi) = C(0)\left[\frac{1-k}{\sqrt{1-2k\,\cos\psi+k^2}} - (1-k)(1-k)k\,\cos\psi\right], \tag{6.8}$$

where $k<1$ is a parameter to be determined, the last equation can be expanded as:

$$C(\psi) = C(0)(1-k)\sum_{n=2}^{\infty} k^n P_n(\cos\psi), \tag{6.9}$$

and by comparing Eqs. (6.7a) and (6.9) one may identify the degree variances as:

$$\sigma_n^2 = C(0)k^n(1-k). \tag{6.10}$$

Moreover, at the correlation length ψ_c one obtains:

$$\frac{1}{2} = \frac{C(\psi_c)}{C(0)} = \frac{1-k}{\Omega} - (1-k) - (1-k)k\,\cos\psi_c, \tag{6.11a}$$

where:

$$\Omega = \sqrt{1-2k\,\cos\psi_c} + k^2. \tag{6.11b}$$

If the correlation length is fixed, the parameter k can be determined iteratively from Eq. (6.11a), e.g. by Newton–Raphson's method:

Iterate:

$$k_{j+1} = k_j - f(k_j)/f'(k_j); \mathrm{j} = 0, 1, 2, \ldots, \tag{6.12a}$$

where:

$$f(k) = \frac{1-k}{\Omega} - (1-k)(1+kt) - \frac{1}{2}, \tag{6.12b}$$

and

$$f'(k) = \frac{(1-t)(1+k)}{\Omega^3} + 1 - t + 2kt, \tag{6.12c}$$

until convergence.

After that k has been determined, and $C(0)$ has been specified, σ_n^2 follows from Eq. (6.10).

Ågren (2004) pointed out that the reciprocal distance covariance model yields too little power to higher degree terms of the covariance model to be realistic. For that reason, this covariance model was combined with the so-called *white noise covariance model* to provide more realistic results for the gravity anomaly-degree variances. The method is further refined and explained in Ågren et al. (2006, 2009a), and we summarize the results next.

It is reasonable to assume that the EGM error-degree variances (dc_n^2), at least in general, increase with degree (n). On the contrary, we should expect that the error-degree variances of the gravity anomalies (σ_n^2) generally decrease with n up to some medium degree, and, for higher degrees, this decay becomes less pronounced. To achieve this result, the error-covariance model for Δg is constructed as a combination of the reciprocal distance model above and a white noise model. The latter model, with variance σ^2, has the degree variances

$$\sigma_n^2 = (2n+1)\frac{\sigma^2}{(M_N+1)^2 - 4};\ n = 2, 3, \ldots M_N, \tag{6.13}$$

where the integer number $M_N = Int\left[\frac{180}{v^\circ}\right]$, v being the block size of the terrestrial gravity data, and, for higher degrees, $\sigma_n^2 = 0$. The combined covariance model is therefore defined such that half of the given σ^2 is delivered from each of the reciprocal and white-noise functions. Assuming that the Nyquist degree $M_N = \pi/$ block size of data and M_{rd} is the assumed upper degree of the reciprocal distance function, this leads to the degree variance model

$$\sigma_n^2 = \begin{cases} 0 & if \quad n < 2 \\ c\mu^n(1-\mu) & if \quad 2 \le n \le M_{rd} \\ \frac{\sigma^2}{2}\frac{(2n+1)}{(M_N+1)^2-4} & if \quad M_{rd} < n \le M_N \\ 0 & if \quad n > M_N \end{cases}, \tag{6.14}$$

where the parameters c and μ of the reciprocal function are now fixed by the equations

$$\sum_{n=2}^{M_{rd}} \left[c\mu^n(1-\mu) - \frac{\sigma^2}{2} \frac{(2n+1)}{(M_N+1)^2 - 4} \right] = \frac{\sigma^2}{2} \tag{6.15}$$

and

$$c\mu^{M_{rd}}(1-\mu) = \frac{\sigma^2}{2} \frac{(2M_{rd}+1)}{(M_N+1)^2 - 4}. \tag{6.16}$$

Here M_{rd} is the maximum degree used in the series for the reciprocal distance-covariance function.

A modified, and to some extent simpler but still efficient, technique for combining the error-degree variances given by Eqs. (6.10) and (6.13), was presented by Ågren et al. (2009a). In the combination of the two covariance models, our experiences have been that the choices of individual covariance function parameters like variance $C(0)$ and correlation length is not so important, but most important is the degree K at which the plots of the spectra of the EGM and the gravity anomaly cross each other. Different parameter choices with the same crossing point K tend to provide very similar modifications (but not the same propagated mean-square error). Hence, the choice of K is that of specifying the upper degree to which the EGM is believed to be better, or at least as good as, the gravity anomalies, and this choice finally specifies the error degree variance model of the terrestrial data by a proper combination of those given by the reciprocal distance and white-noise models. In practice, this means that one tentatively adds the weighted sum of error degree variances of the two models and compares for the resulting K and geoid-height standard errors in the LSMSA technique, as well as RMS differences, to available GNSS-levelling geoid heights, etc.

6.5.2 Which EGM Should Be Used?

There is a variety of EGMs available for combination with terrestrial gravity data. The question is how to select the best one for the region to be studied. A simple procedure would be to compare the EGM-derived geoid heights with those from GNSS/levelling of the region. However, in this context, one should also consider the maximum degree (M) of the EGM to be used in the combination. In general, the larger M, the better is the agreement. The most important criterion for this comparison is the RMS difference between the two types of geoid heights.

A better result is obtained by putting different EGMs into the LSMSA procedure for comparison against GNSS-levelling, and the more complete the computational procedure is, the more conclusive will be the outcome of the comparison.

However, in many developing countries, there are no or, at least, not a sufficient number of GNSS/levelling heights available for this type of pre-study. In such cases, the comparison of the EGM with available terrestrial gravity anomalies could be useful. However, in such a comparison, one must bear in mind that the EGM does not contain the fine spectrum available in the point gravity data. Ulotu (2009) conducted such an analysis for Tanzania by smoothing the terrestrial gravity data by space-shuttle-derived topographic data.

6.5.3 Choice of Cap Size

In the ideal situation with only random errors with expectation zero, there is no optimum cap size for the numerical integration by the LSMSA method, but the cap size should only be governed by the area of gravity data available. This is because the least squares method automatically includes all data in an optimum way. In practice, due to systematic errors, this is not the case, but a limited cap size should be chosen to reduce the influence of systematic/gross errors in the gravity data. On the other hand, a too small cap size should not be chosen, which would impair the solution due to the rejected data outside the cap.

A standard way of choosing cap size is to make some preliminary LSMSA estimates and compare them with GNSS/levelling geoid heights. In this way, the cap size can easily be tuned towards the best agreement of the two independent geoid estimators. However, one should keep in mind that this technique forces the gravimetric LSMSA solution towards the GNSS/levelling solution, which could be, and of course also is, in error. If the LSMSA geoid model will be used together with GNSS data as a corrective surface for height determination, this technique should be used in full, but if the goal is to determine a geoid model, the systematic impact from the errors of the GNSS/levelling data should be considered.

6.5.4 Numerical Considerations in Determining Modification Parameters

The linear system of equations for solving the least squares modification parameters s_n is ill-conditioned (Ellmann 2005a). Although the system is badly conditioned, this feature is harmless to the result. Ågren (2004) explained the bad conditioning of the system as caused by the behaviour of the modified Stokes' formula in the remote-zone (where it is not used in the modified Stokes' formula): Many sets of modification parameters agree well in the near-zone and disagree in the remote-zone (where they are harmless). In practice, the system of Eq. (3.28a) can be solved by some stabilization method like Tikhonov regularization or Singular Value Decomposition. The latter technique is used in the KTH software. See also publications by Ellmann (2005a, b, 2012) and related software.

6.5.5 Comparison of DWC in LSMSA and Other Methods

In the LSMSA technique, the downward continuation of the gravity anomaly to sea level is directly applied to the effect on the geoid height as an additive constant. In this way, this computational step becomes more stable compared with the technique used in RCR, where the gravity anomaly is directly continued downward by inversely solving Poisson's integral formula. Another, but approximate, method is to make the dwc by limiting it to the first term in a Taylor series of the gravity anomaly. Here, the gradient can be computed by an integral formula with gravity anomalies.

Practical experiments in solving Poisson's integral (Martinec 1998, Chap. 8) revealed that gravity anomalies at cap sizes down to 5′ (but not for more detailed data) in the Canadian Rocky Mountains could be handled with stable solutions.

Ågren (2004, Sects. 5.3–4), based on synthetic gravity-field models, numerically compared the techniques for downward continuation used in the LSMSA method and RCR-technique, as well as the gradient method. He came to the conclusions that "Sjöberg's method is superior to using the gradient method", and "this method is considerably more accurate than the inversion by Poisson's equation".

6.6 Case Studies

In Table 6.1, we summarize a number of studies and/or applications of the KTH method for geoid or quasigeoid determinations.

6.6.1 Comparisons of Methods

Some of the papers listed in Table 6.1 include comparison of the LSMSA/KTH technique with other methods, primarily with the RCR method. A few of these comparisons will be summarized here. For other comparisons the reader is referred to the references.

Ellmann (2004, 2005a, b) compared several deterministic and three stochastic modifications of Stokes' formula in the Baltic countries, and it was concluded that the KTH approach was the preferred method.

On behalf of the International Geoid Service, the Auvergne test area was established in France, consisting of about 240,000 gravity data points covering an area of 6° × 8°; a digital elevation model based on three Shuttle Radar Topography Mission height data, covering an area of 8° × 10°, and 75 GPS-levelling data points, all at 1st order levelling data points or with connections to such points. The accuracy of the GPS-levelling data is estimated to 2 cm or slightly better (Duquenne 2007).

Table 6.1 Summary of regional studies and/or applications of the LSMSA/KTH method

Author	Country/region	Geoid/Q-geoid	Remarks
Fan (1989)	Theory	–	theoretical study
Nsombo (1996)	Zambia	geoid	spectr. comb.; RMS = 70
Nahavandchi (1998)	Sweden	Q-geoid	RMS = 5.5
Hunegnaw (2001)	Ethiopia	geoid	Tests for dwc. RMS = 46
Ellmann (2004)	Baltic region	Geoid	RMS fits of 2.8, 5.6 and 4.2 cm for Estonia, Latvia and Lithuania
Ågren (2004)	Theory	Geoid/Q-geoid	error degree variances, dwc, etc.
Kiamehr (2006)	Iran	Geoid	RMS = 29
Daras (2008)	Greece	–	–
Ågren et al. (2009a)	Sweden	Q-geoid	σ = 20, 32 for KTH and RCR
Ågren et al. (2009b)	Auvergne, France	Q-geoid	Comparison of seven methods
Abdallah (2009)	Sudan	Geoid	–
Ulotu (2009)	Tanzania	Geoid	σ = 28
Abbak et al. (2012)	Konya, Turkey	Geoid	Comparison KTH and RCR
Yildiz et al. (2012)	Auvergne, France	–	Comparison KTH, RCR, LSC
Danila (2012)	Molodova	Q-geoid	–
Isik and Erol (2015a)	Turkey	Geoid	σ = 14.8, 16.6, 29.2 for KTH, RCR, EGM
Isik and Erol (2015b)	NW Turkey	Geoid	σ = 9.5/6.9 for KTH/RCR after corrector surface fit
Abdalla and Mogren (2015)	S. Arabia	Geoid	
Ssengendo (2015)	Uganda	Geoid and Q-geoid	σ = 8.4 and 7.3
Ågren et al. (2015) and (2016)	Fennoscandia	Quasigeoid	σ = 3 for whole area = 1–2 in flatter countries

Some studies compare with the RCR method. RMS = fit with GNSS or Doppler in cm; σ = estimated standard error after fitting (cm)

Table 6.2 Results from the international software comparison of quasigeoid estimation at Auvergne test area (Ågren et al. 2009b)

	KTH	PoliMi	IFE	NBI	DGS	IGN
RMS	29	36	35	67	35	37
σ_ζ	22.7	31.2	30	64.5	30	32.3

RMS = RMS fit with GPS/levelling after 3-parameter reduction and σ_ζ = RMS gravimetric Q-geoid accuracy at the 75 GPS-levelling points with an estimated accuracy of 32 mm. Unit: mm

Table 6.3 Comparisons of 4-parameter fits of quasigeoid determinations using gravimetric methods versus GPS/levelling at the Auvergne test area (from Yildish et al. 2012)

Gravimetric models	Max	Min	RMS
RCR−	68	−118	39
RCR+	58	−67	29
KTH	95	−51	24
LSC	64	−65	31
EGM2008	138	−68	29

RCR− and RCR+ = RCR without and with Stokes' kernel modification. LSC = Least Squares Collocation. Unit: mm

A joint international comparison of six techniques was reported at the Hotine-Marussi symposium in Rome 2009 (Ågren et al. 2009b). Besides the KTH method, the test included modified Stokes' method with spectral combination (IFE, Hannover), fast collocation approach (PoliMi; Politecnico di Milano), least squares collocation (NBI, University of Copenhagen), 1-D spherical FFT (DGS; University of Thessaloniki) and Stokes' integration (IGN, France). In Table 6.2, the RMS fits (after removal of a 3-parameter systematic-error model of bias and two slopes) for six of the involved techniques and GPS-levelling are shown on the first row, ranging between 29 and 67 mm. The second row shows the estimated standard errors of the geoid estimates from the included methods, obtained after removing the estimated standard errors of 25 and 20 mm squared for GPS and levelling observations, respectively, from the mean square differences on the first row. The result shows that the KTH method performs significantly better than other methods.

The above test did not include the Danish/Copenhagen school versions of the RCR and collocation approaches. A second test with these methods was conducted and reported in Yildiz et al. (2012); see Table 6.3. The RMS fit using a 4-parameter model for systematic errors yielded 2.4 cm for the KTH method, while other methods fit to 2.9 cm or slightly worse. As emphasized by Sjöberg (2005), the study shows that the RCR− (without the modification of Stokes function) performs less well than RCR+ (with modification). It is also interesting that EGM 2008 agreed well with most other methods, indicating that it is based on good data in the region.

- In Ågren et al. (2009a), the KTH and RCR techniques were compared with GPS-levelling over Sweden. The RCR technique is that practiced in the previous Nordic Geodetic Commission (NKG) geoid model from 2002, but the data has been updated. The results of the comparison are reported in Fig. 6.1. The 1-parameter RMS fits are 22 and 40 mm for KTH and NKG models, while these values reduce to 20 and 32 mm in the 4-parameter fits. It is obvious from these numbers and Fig. 6.1 that the NKG model includes long-wavelength NW-SE systematic errors after 1-parameter fit, which are mostly removed after the 4-parameter fit, while the results in the comparison with the KTH method does not suffer from this error (implying that the error cannot be referred to GPS-levelling). Therefore, one may conclude:

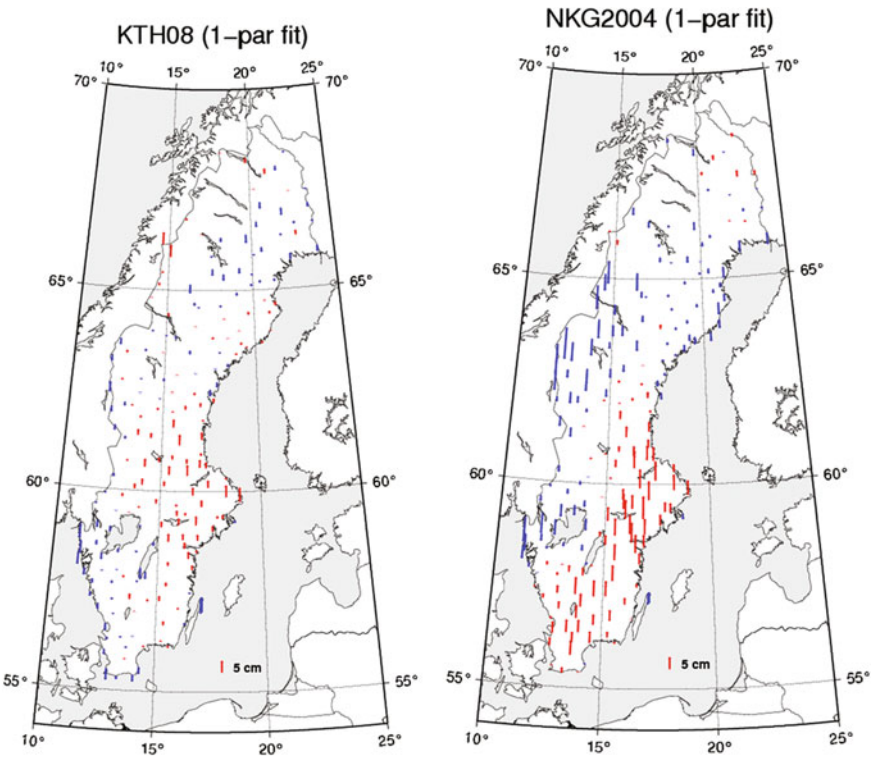

Fig. 6.1 196 GPS-levelling residuals for the KTH and NKG models after 1-parameter fits. (The scales are given by the 5 cm arrows to the south-east; KTH08 is based on LSMSA, and NKG2004 uses RCR-method) (Ågren et al. 2009a)

- As the residuals of the LSMSA versus GPS-levelling are small, the larger residuals for NKG are likely to be caused by (systematic) errors in the NKG technique rather than in GPS levelling.
- There are obvious long-wavelength errors in NKG residuals.
- With estimated accuracies of 15 and 7.5 mm of the GPS and levelling observations, respectively, the 1- and 4-parameter RMS fits correspond to standard-error estimates of 16 and 11 mm, respectively, in the LSMSA models.

Abbak et al. (2012) conducted another comparison of the KTH and RCR methods versus GPS-levelling. The test area was located in the mountainous area around Konya in central Turkey, limited by latitudes 37–39N and longitudes 31.5–35E, an area with a very limited amount of terrestrial gravity data. For the evaluation, 20 GPS-levelling data points were used with an estimated accuracy of 3.6 cm. The resulting RMS fits, using a 7-parameter model to absorb systematic deviations between models, were 6.7 and 9.8 cm for the KTH and RCR methods, respectively, corresponding to gravimetric geoid model accuracies of 5.7 and

9.1 cm, respectively. The two models disagree by more than 1 m in the SW corner of the test area, and the reason for that has not been clarified.

6.6.2 NKG Quasigeoid Model 2015 (NKG2015 Geoid)

Next, we present some details and results on the NKG 2015 quasigeoid model determination and results [courtesy J. Ågren, Lantmäteriet; Ågren et al. (2015 and 2016)]. The model was computed by J. Ågren as a result of the efforts of the NKG working group on geoid and height determination.

- Method:
  - Least Squares Modification of Stokes' formula (stochastic kernel modification) with Additive corrections (LSMSA- or KTH-method).
  - Implementation written by J. Ågren LM/KTH (most key programs the same as in KTH Geolab software package)
  - Stokes' integration: 2° spherical cap using the 1-D-FFT method (100% zero padding)
  - EGM: GO_CONS_GCF_2_DIR_R5 with M = 300
  - Modification: **Unbiased** Least Squares Modification with gravity anomaly error degree variances (DV) chosen as a combination of band-limited white noise and the reciprocal distance model (1 mGal white noise, 0.5 mGal correlated noise with correlation length 0.25°), formal EGM error DV and re-scaled Tscherning and Rapp signal DV.
- Gravity anomaly gridding:
  - Remove-interpolate-restore using **Least Squares Collocation** (GEOGRID) on the gravity anomaly **reduced** for:
    - EGM effect by GO_CONS_GCF_2_DIR_R5 with M = 240.
    - RTM effect computed by terrain corrections only with 3″ × 6″ to 25 km (in flat areas 0.0025 × 0.0050°) and 0.01 × 0.02° to 167 km. Restore 0.01 × 0.02°.
    - Ice model effect (i.e. effect due to the density difference between rock and ice) computed by rectangular prisms (close to the point) and spherical integration (further away) using the Airy-Heiskanen isostatic compensation method.
- Other comments:
  - EIGEN-6C4 with M = 2190 was used in areas without gravity observations to generate pseudo-observations.
  - Cross validation on the reduced gravity anomaly with 20 mGal rejection limit.

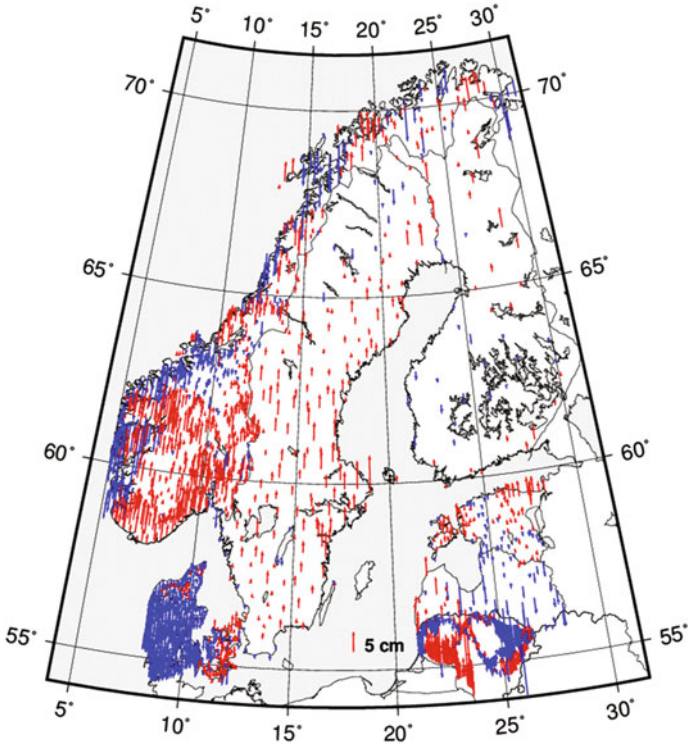

Fig. 6.2 Differences GNSS-levelling minus NKG2015 quasigeoid model after a common 1-parameter bias fit using the national EVRS realizations

Table 6.4 Statistics for the GNSS/levelling minus NKG2015 quasigeoid models after a common 1-parameter fit for the whole area using the national EVRS realizations. Unit: metre

(metre)	No. of points	Before fit	After common 1-parameter fit		
	#	Mean	Mean	Std	RMS
All	**2538**	**−0.4876**	**0.0000**	**0.0299**	**0.0298**
Denmark	675		−0.0166	0.0169	0.0237
Estonia	114		0.0071	0.0147	0.0163
Finland	50		0.0052	0.0214	0.0307
Latvia	54		−0.0186	0.0247	0.0307
Lithuania	546		0.0021	0.0332	0.0333
Norway	902		0.0070	0.0333	0.0340
Sweden	197		0.0180	0.0185	0.0258

Table 6.5 Approximate GNSS/levelling accuracy (from J. Ågren; private communication) and derived accuracy of the NKG2015 geoid model. Unit: mm

Country	Denmark	Estonia	Finland	Latvia	Lithuania	Norway	Sweden
GNSS/ levelling	15	15	15	20	20	15	15
NKG2015	8	–	15	15	27	30	11

- The quasigeoid was computed corresponding to the new IHRS W_0 value.
- A small-land uplift correction was applied for the fact that the "epoch" of the EGM differs from the epoch 2000.0 has been applied (NKG2005LU). (The latter is very small, below 4 mm.)

The results are demonstrated in Fig. 6.2 and Tables 6.4 and 6.5. The RMS agreement between GNSS/levelling and NKG2015 geoid models all over the area is 30 mm with national mean differences varying from −19 mm (Latvia) to +18 mm (Sweden). The national standard errors of differences vary between 14 mm (Estonia) and 33 mm (Norway and Lithuania).

By considering the standard deviation differences between GNSS/levelling and NKG2015 geoid model, as well as the approximate GNSS/levelling accuracies, the estimated accuracies by country of NKG2015 geoid model is obtained as reported in Table 6.5. The estimated accuracy is 3 cm or less and reaches 1 cm for Denmark and Sweden. For more details see Ågren (2016).

Ssengendo (2015) studied the LSMSA technique for the proposal of a new vertical datum in Uganda. As there were a very limited number of existing first-order vertical control points in the country, a new height network was requested. As is the case in many (not only developing) countries, the old height network in Uganda is based on normal-orthometric heights. Ssengendo used these heights for GPS/levelling of geoid or possibly quasigeoid heights, which was in doubt as the levelled heights were normal-orthometric heights. Comparing with gravimetric quasigeoid and geoid heights determined by the LSMSA method, he found RMS differences with GNSS-levelling results of 8 and 15 cm, respectively, after 4-parameter fits, suggesting that the normal-orthometric heights can be regarded as approximations to normal heights rather than to orthometric heights. Based on this finding, Ssengendo (ibid.) proposed that the future vertical datum of Uganda should be based on normal heights and a quasigeoid model.

6.7 Concluding Remarks

In theory, the RCR and LSMSA methods should provide the same geoid and quasigeoid estimates, but, in numerical practice, they usually differ. To some extent, the differences are due to numerical interpolations and integrations, but there are also differences related to the specific methods as described above in Sect. 6.4.

In RCR applications, it is common to select a suitable topographic reduction model (e.g. Helmert condensation, isostatic reduction or Residual Terrain Modelling) to make the gravity anomaly smooth and suitable for interpolation and integration as well as the PITE small. The choice of reduction model is not relevant in the LSMSA method as it computes the combined topographic effect on the geoid height, which cannot contain any topographic reduction. Also, in quasigeoid determination, the combined topographic effect is zero.

As shown in the previous section, the estimated accuracies of the quasigeoid heights in Denmark and Sweden are of the order of 1 cm. Along with future improvements in GNSS position, the user will ask for even better geoid models, and it is not unlikely that the 5 mm geoid will be required. Therefore, the NKG started a project to investigate the requirements for such a model. This model was defined as a quasigeoid model with an overall relative accuracy (standard error) of 5 mm, and there is no corrector surface applied based on GNSS-levelling. So far it has been shown that requirements will be fulfilled for Sweden (Ågren and Sjöberg 2014; Sjöberg and Ågren 2014):

- The average 5 km resolution and present accuracy of the gravity net will be sufficient provided that the data are updated for some systematic errors and data gaps, and the data in surrounding areas, e.g. the Baltic Sea, need improvements.
- Systematic errors in the Digital Elevation Models (DEMs) are not a problem over Sweden, where a high-resolution and quality DEM is available.

References

Abbak RA, Erol B, Ustun A (2012) Comparison of the KTH and remove-compute-restore techniques to geoid modelling in a mountainous area. Comput Geo Geosci 48:31–40

Abdalla A (2009) Determination of a gravimetric geoid model of Sudan using the KTH method. (pdf 3,1 MB). Masters of Science Thesis in Geodesy No. 3109, TRITA-GIT EX 09-001. Supervisor: Huaan Fan. Jan 2009

Abdalla A, Mogren S (2015) Implementation of a rigorous least-squares modification of Stokes formula to compute a precise geoid model over Saudi Arabia (SAGEO13). Can J Earth Sci. doi:10.1139/cjes-2014-0192

Ågren J (2004) Regional geoid determination methods for the era of satellite gradiometry. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden

Ågren J (2016) NKG2015 geoid model project—Compilation and validation of the computation centre solutions. Internal LM project report, Gävle, Sweden

Ågren J, Kiamehr R, Sjöberg LE (2006) The Swedish geoid as evaluated by the method of least-squares modification with additive corrections. Paper presented to the 1st symposium of the International gravity field service, Istanbul, 28 Aug–1 Sept

Ågren J, Sjöberg LE (2014) Investigations of the requirements for a future 5 mm quasigeoid model over Sweden. In Marti U (ed) Gravity, geoid and height systems, international association of geodesy symposia 141. doi:10.1007/978-3-319-10837-7_18

Ågren J, Sjöberg LE, Kiamehr R (2009a) The new gravimetric quasigeoid model KTH08 over Sweden. J. Appl Geod 3(3):143–153

Ågren J, Barzaghi R, Carrion D, Denker H, Duquenne H, Grigoriadis VN, Kiamehr R, Sona G, Tscherning CC, Tziavos IN (2009b) Different geoid computation methods applied on a test dataset: results and considerations. Poster presented at Hotine-Marussi symposium, Rome, p 2009

Ågren J, Strykowski G, Bilker-Koivula M, Omang O, Märdla S, Oja T, Liepiņš I, Paršeliūnas E, Forsberg R, Kaminskis J, Ellmann A, Sjöberg LE, Valsson V (2015) On the development of the new Nordic gravimetric geoid model NKG2015. Presentation at 26th IUGG general assembly, Prague, Czech Republic, June 22–July 2, 2015

Ågren J, Strykowski G, Bilker-Koivula M, Omang O, Märdla S, Oja T, Liepiņš I, Paršeliūnas E, Forsberg R, Kaminskis J, Ellmann A, Sjöberg LE, Valsson V (2016) The NKG2015 gravimetric geoid model for the Nordic-Baltic region. Presentation at the international symposium on GGGH systems, 19–23 Sept Thessaloniki, Greece

Danila U (2012) Mold2012—a new gravimetric quasigeoid model over Moldova. Licentiate Thesis in Geodesy, Royal Institute of Technology, Stockholm, Sweden

Daras I (2008) Determination of a gravimetric geoid model of Greece using the method of KTH. TRITA-GIT EX 08-02. Supervisor: Huaan Fan. January 2008

Duquenne H (2007) A data set to test geoid computation methods. Harita Dergisi. Special Issue 18:61–65

Ellmann A (2004) The geoid for the Baltic countries determined by the least squares modification of Stokes' formula. Doctoral Thesis in Geodesy, Royal Institute of Technology, Stockholm, Sweden

Ellmann A (2005a) Computation of three stochastic modifications of Stokes's formula for regional geoid determination. Comput Geosci 31(6):742–755

Ellmann A (2005b) Two deterministic and three stochastic modifications of Stokes's formula: a case study for the Baltic countries. J Geodesy 79:11–23

Ellmann A (2012) Using high-resolution spectral models of gravity anomaly for computing stochastic modifications of Stokes's formula. Comput Geosci 39. doi:10.1016/j.cageo.2011.07.001

Ellmann A, Vaníček P (2007) UNB application of Stokes'-Helmert's approach to geoid computation. J Geodyn 43:200–213

Fan H (1989) Geoid Determination by global geopotential models and integral formulas. Ph.D. Thesis, Division of Geodesy, Royal Institute of Technology, Stockholm, Sweden

Hunegnaw A (2001) Geoid Determination over Ethiopia with emphasis on downward continuation of gravity anomalies

Kiamehr R (2006) Precise gravimetric geoid model for Iran based on GRACE and SRTM data and the least-squares modification of stokes' formula: with some geodynamic interpretations (Ph.D. Thesis)

Isik M, Erol B (2015a) Geoid modelling in Turkey using remove-compute-restore and least squares modification of Stokes' integral methods. IUGG General Assembly, Prague, Check Republic, June 2015

Isik M, Erol B (2015b) Comparison of Remove-Compute-restore and KTH methods for gravimetric geoid modelling in Northwest of Turkey. Presented to the XXVth International symposium on modern technologies, education and professional practice in geodesy and related fields, Sofia, Bulgaria 6–7 Nov 2015

Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture Notes in Earth Sciences No. 73, Springer

Nsombo P (1996) Geoid determination over Zambia. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden

Nahavandchi H (1998) Precise GPS-gravimetric geoid determination with improved topographic corrections over Sweden. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden

Sjöberg LE (1986) Comparison of some methods of modifying Stokes' formula. Bollettino di Geodesia E Scienze Affini XLV(3) 46(2):229–248

Sjöberg LE (2003) A computational scheme to model the geoid by the modified Stokes' formula without gravity reductions. J Geodesy 77:423–432

Sjöberg LE (2005) A discussion on the approximations made in the practical implementation of the remove-compute-restore technique in regional geoid modeling. J Geod 78:645–653. doi:10.1007/s00190-004-0430-1

Sjöberg LE (2009) The terrain correction in gravimetric geoid determination—is it needed? Geophys J Int 176:14–18

Sjöberg LE, Ågren J (2014) Investigations for the requirements of a 5 mm geoid model- a project status report. Poster presentation at the NKG general assembly, 1–4 Sept 2014, Gothenburg, Sweden

Ssengendo R (2015) A height datum for Uganda based on a gravimetric quasigeoid model and GNSS/levelling. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden

Ulotu P (2009) Geoid model of Tanzania from sparse and varying gravity data density by the KTH method. Ph.D. thesis, Royal Institute of Technology, Stockholm, Sweden

Vanicek P, Kleusberg A (1987) The Canadian geoid ± Stokesian approach. Manuscr Geod 12:86–98

Wong L, Gore R (1969) Accuracy of geoid heights from modified Stokes kernels. Geophys J R Astr Soc 18:81–91

Yildiz H, Forsberg R, Ågren J, Tscherning CC, Sjöberg LE (2012) Comparison of remove-compute-restore and least squares modification of Stokes' formula techniques to quasigeoid determination over the Auvergne test area. J Geod Sci 2:53–64

Chapter 7
Further Tools in Physical Geodesy

Abstract This chapter deals with a variety of questions in physical geodesy. Various methods to determine the quasigeoid, such as Molodensky's and Bjerhammar's methods as well as Stokes' extended formula at point level, are briefly presented. It also deals with the correction from the height anomaly to the geoid height, the combination of geoid models from gravimetric and geometric data, some methods to determine the potential at the geoid (W_0), which is needed for absolute geoid determination, and spectral smoothing and combinations. Finally an emerging relativistic technique for direct-height and geoid-height determinations is concisely presented.

Keywords Atomic clocks · Bjerhammar's method · Molodensky's method · Potential at the geoid · Quasigeoid

7.1 Quasigeoid Determination

As briefly introduced at the beginning of Chap. 4, the concepts of *the quasigeoid height* (or *height anomaly*) and the *surface-gravity anomaly* were created by Molodensky et al. (1962). The surface connecting all points Q with normal potential $U_Q = W_P$ (with the geopotential W located at the surface point P) is called the *telluroid*, and its projection on the reference ellipsoid is denoted the quasigeoid (see Fig. 4.3). The great advantage of introducing the quasigeoid as a reference surface instead of the geoid is that it can be determined without any information about the topographic density distribution.

A few methods to determine the quasigeoid are presented in the next sections. Two additional methods (the RCR and LSMSA techniques) were already presented in Sect. 6.3.

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_7

7.1.1 Molodensky's Method

As a contrast to Stokes' integration of the geoid height, Molodensky presented a solution to the height anomaly that integrates surface-gravity anomalies over the Earth's surface. The method utilizes a surface density (κ) integral on the Earth's surface (E) of the disturbing potential:

$$T = \iint_E \frac{\kappa}{l} dE, \tag{7.1a}$$

where:

$$dE = r^2 \frac{d\sigma}{\cos \beta}. \tag{7.1b}$$

Here r is the radius vector of the integration point and β is the inclination angle of the terrain. By inserting (7.1a) into the boundary condition in spherical approximation, (3.23b), one obtains an observation equation for the surface gravity anomaly:

$$2\pi\kappa_P - \iint_\sigma \left(\frac{3}{2l} + \frac{r^2 - r_P^2}{2l^3}\right) \frac{r^2}{r_P} \kappa \frac{d\sigma}{\cos \beta} = \Delta g_P. \tag{7.2}$$

This is a linear Fredholm integral equation of the second kind that numerically can be written as the matrix equation

$$\boldsymbol{\kappa} + \mathbf{K}\boldsymbol{\kappa} = \boldsymbol{\Delta g}, \tag{7.3a}$$

which possibly can be solved by iteration:

$$\boldsymbol{\kappa}^{(k+1)} = \boldsymbol{\Delta g} - \mathbf{K}\boldsymbol{\kappa}^{(k)}; \quad k = 0, 1, 2, \ldots . \tag{7.3b}$$

By inserting the solution for the density anomaly into (7.1a) and applying Bruns' formula, the height anomaly is obtained:

$$\zeta_P^{(k)} = \frac{1}{\gamma_Q} \iint_E \frac{\kappa^{(k)}}{l} dE. \tag{7.4}$$

Molodensky developed the iteration procedure further by successive approximations of the form (Heiskanen and Moritz 1967, Sect. 8.7):

$$\zeta = \zeta_0 + \zeta_1 + \ldots = \frac{R}{4\pi\gamma_Q}\iint_\sigma S(\psi)G_0 d\sigma + \frac{R}{4\pi\gamma_Q}\iint_\sigma S(\psi)G_1 d\sigma + \ldots, \quad (7.5a)$$

where: $G_0 = \Delta g$ and

$$G_1 = \frac{R^2}{2\pi}\iint_\sigma \frac{h - h_P}{l_0^3}\left(G_0 + \frac{3\gamma_0}{2R}\zeta_0\right)d\sigma. \quad (7.5b)$$

Iteration to higher-order terms can be found in Moritz (1980, pp. 362–364).

Although the solution is elegant, it suffers from the cumbersome integration on the rough surface of the Earth. In fact, this is only possible if one assumes/approximates the Earth as a star shaped body with slopes not exceeding 45°. According to Moritz (1980, Sect. 47), the series is an asymptotic series for larger slopes, implying that the series provides a good approximation when truncated to a few terms only, but will be impaired when including more terms. However, in rough terrain the slope of the terrain by far exceeds 45°, and the approximation is too crude for the purpose, which aims at determining the (quasi) geoid to, say, 1 cm.

7.1.2 Bjerhammar's Method and Collocation

Krarup (1969) emphasized that traditional geodesists derived the geoid estimates from the perspective of solving a boundary value problem. This view is also at hand in Molodensky's approach to determine the quasigeoid. As a contrast, Bjerhammar (1962) and (1963) developed a method based on solving a linear system of discrete observation equations. The method includes *analytical continuation* of the gravity anomaly. Using Poisson's integral formula for the observed surface-gravity anomaly

$$\Delta g_P = \frac{R^2\left(r_P^2 - R^2\right)}{4\pi r_P}\iint_\sigma \frac{\Delta g^*}{l_P^3}d\sigma, \quad (7.6)$$

an integral equation is formed for the fictitious gravity anomalies Δg^* on the internal sphere (the Bjerhammar sphere) of radius R (see Fig. 7.1). Once Δg^* has been solved, the classical forward solutions of Eq. (7.6)

$$T_P = \frac{R}{4\pi}\iint_\sigma S(r_P, \psi)\Delta g^* d\sigma \quad (7.7)$$

and similar expressions for the external Vening Meinesz' formulas (Eq. 3.44), etc., can be utilized for solving the gravity anomaly, disturbing potential and deflections

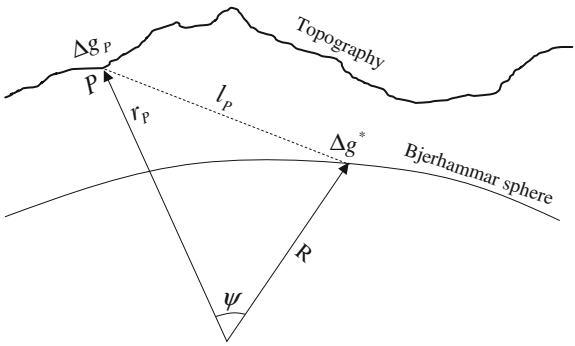

Fig. 7.1 Reduction of measured gravity anomaly to the internal sphere (Bjerhammar sphere)

of the vertical at any point on or outside the Earth's surface. Equation (7.7) inserted into Bruns' formula yields the height anomaly.

This is the simple formulation of Bjerhammar's method. However, its practical implementation is not easy. First of all, it is not likely that the fictitious gravity anomaly exists in the continuous case. [It is said that also Molodensky, at an early stage, was considering a similar technique, but gave it up.] However, Bjerhammar argued that the practical case will always deal with a finite number of observations, which always has a solution. Then Eq. (7.6) leads to the matrix equation

$$\mathbf{K}\Delta\mathbf{g}^* = \Delta\mathbf{g}. \tag{7.8}$$

As this equation is a first-order Fredholm integral equation, it quickly becomes numerically unstable for an increasing number of observations, unknowns and elevation, but there is always a unique solution available, e.g. by utilizing the method of Singular Value Decomposition (e.g. Bjerhammar 1973, pp. 345–347).

If the system (7.8) is over-determined (with more observations than unknowns), formally there is a unique least squares solution (see section 2.1.1)

$$\Delta\hat{\mathbf{g}}^* = \left(\mathbf{K}^T\mathbf{P}\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{P}\Delta\mathbf{g} \tag{7.9a}$$

that minimizes the sum of squares of residuals $\boldsymbol{\varepsilon}^T\mathbf{P}\boldsymbol{\varepsilon}$, where $\mathbf{P}$ is a weight matrix.

If the system is under-determined, there is a unique least squares solution that minimizes the norm $(\Delta g^*)^T\,\mathbf{Q}^{-1}\Delta g^*$:

$$\Delta\hat{\mathbf{g}}^* = \mathbf{Q}\mathbf{K}\left(\mathbf{K}\mathbf{Q}\mathbf{K}^T\right)^{-1}\Delta\mathbf{g}. \tag{7.9b}$$

Inserting the solution of Eq. (7.9b) into (7.8), one gets back to the surface observations (which are not fictitious). This procedure can be generalized to the prediction that other gravity anomaly related parameters justby changing the matrix $\mathbf{K}$ to a suitable discretized integral (see Eq. 7.10 below). This means that this technique yields fictitious (biased) predictions inside the masses and unbiased predictions on the surface and outside the masses.

As shown by Sjöberg (1975) and (1978), this solution leads to the solution by collocation when choosing the norm matrix **Q** as a diagonal matrix with diagonal elements:

$$Q_{ii}^{-1} = \Delta S_i / S, \text{ where } \sum_{i=1}^{N} \Delta S_i = S = 4\pi R^2, \tag{7.9c}$$

and ΔS_i is a surface element on the sphere of radius R. To see this, let a quantity u be linearly related with the gravity anomaly by the coefficient vector **k**. Then its prediction becomes:

$$\hat{u} = \mathbf{k}^T \Delta \mathbf{g}^* = \mathbf{k}^T \mathbf{Q} \mathbf{K}^T \left(\mathbf{K} \mathbf{Q} \mathbf{K}^T \right)^{-1} \Delta \mathbf{g}, \tag{7.10}$$

and when the number of surface elements (N) on the Bjerhammar sphere (as well as observations) goes to infinity, it follows that:

$$Q_{ii}^{-1} \to \frac{dS}{S} \text{ and } \hat{u} \to \mathbf{c}^T \mathbf{C}^{-1} \Delta \mathbf{g}, \tag{7.11}$$

where (in the concept of collocation) **c** is the cross-covariance matrix between u and $\Delta \mathbf{g}$ and **C** is the auto-covariance matrix of the gravity anomaly with elements

$$c_{ji} = \lim \left(\mathbf{k}^T \mathbf{Q} \mathbf{K}^T \right)_{ji} \text{ and } \mathrm{C}_{ij} = \lim \left(\mathbf{K} \mathbf{Q} \mathbf{K}^T \right)_{ij} \text{ as } N \to \infty \text{ and } (\Delta S_i)_{\max} \to 0. \tag{7.12}$$

Bjerhammar's method can also be seen as a parallel to the mass-point technique used by many scientists for solving problems in physical geodesy. These types of methods often use Newton's formula or similar ones for predefined locations of point masses, while Bjerhammar uses Poisson's formula for gravity anomalies. If all mass points are located on the Bjerhammar sphere, the mass point technique is very similar to Bjerhammar's Dirac approach (Bjerhammar 1975) to solve his problem. Sjöberg (1975 and 1978) generalized Bjerhammar's method by using various fictitious layers on the inner sphere.

Bjerhammar's method was first met with scepticism among geodesists, and their main argument was that Δg^* does not exist, is not unique and/or cannot be determined to sufficient accuracy. Bjerhammar agreed that this was true in the continuous case, but he argued that in the practical case there is only a finite number of observations that can provide a unique solution to Δg^* . Among the first prominent geodesists to appreciate Bjerhammar's arguments were H. Moritz, who devoted one section of Heiskanen and Moritz (1967, Sect. 8.7) to this idea, and another believer was M. Hotine (Hotine 1969, pp. 323–324). Later (Moritz 1980, Sect. 45) developed the method of analytical continuation further, but now with direct application at point level and not at the Bjerhammar sphere (see Sect. 7.1.3).

T. Krarup thought that Bjerhammar's arguments were far from convincing, and he had in mind to prove that Bjerhammar's technique does not work in practice.

However, while studying the problem, he became fascinated, and he developed the method further to least-squares collocation (Krarup 1969). This method is also a generalization of the statistical least-squares prediction presented by Heiskanen and Moritz (1967, Chap. 7). While the prediction technique is restricted to interpolation and extrapolation, collocation can be used, e.g. in estimating the disturbing potential, the height anomaly and deflections of the vertical from surface-gravity anomalies. (Section 2.2.1.) Importantly, Krarup's report also included a theoretical proof of convergence of the approximation by this method (the Runge-Krarup theorem). In this way, both Bjerhammar's method and collocation have a foundation in modern approximation theory.

Collocation, as presented by Krarup (1969), is based on a statistical concept with covariance functions developed by Moritz (1972 and 1980). However, the conversion from Bjerhammar's deterministic approach to the statistical concept is not at all evident and must be regarded with scepticism. Lauritzen (1973) proved that the empirical covariance function on the sphere is not ergodic for a normally distributed gravity field, implying that it is not possible to determine the degree variances of the isotropic covariance function even if the gravity field were known all over the sphere. As a result, collocation should first of all be regarded as an approximation technique that works best for interpolation tasks and to some extent for preliminary generalized interpolation, which can hardly reach the highest accuracy. Nevertheless, collocation is much applied also today in physical geodesy (e.g. Tscherning 2013).

7.1.3 Analytical Continuation at Point Level

The quasigeoid can also be determined by Stokes' formula similar to the geoid height by applying Bruns' formula to Eq. (7.7) as in Bjerhammar (1962):

$$\zeta_P = \frac{R}{4\pi\gamma} \iint_\sigma S(r_P, \psi) \Delta g^* d\sigma, \tag{7.13a}$$

where γ is normal gravity at normal height, and $S(r_P, \psi)$ is the extended Stokes' formula (Heiskanen and Moritz 1967, p. 233):

$$S(r_P, \psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n-1} \left(\frac{R}{r_P}\right)^{n+1} P_n(\cos\psi), \tag{7.13b}$$

which was presented in a closed form in Eq. (3.33c).

As an attractive alternative, the height anomaly can also be determined at computational point level by the original Stokes' formula by analytically continued gravity anomalies $\Delta g(r_P, Q)$ on a sphere through the computation point P:

$$\zeta_P = \frac{r_P}{4\pi\gamma} \iint_{\sigma} S(\psi)\Delta g(r_P, Q) d\sigma. \tag{7.13c}$$

From the previous formula, one can derive the following convenient form for computing the height anomaly:

$$\zeta_P = (\zeta_P) + d\zeta_{dwc}, \tag{7.14a}$$

where:

$$(\zeta_P) = \frac{r_P}{4\pi\gamma} \iint_{\sigma} S(\psi)\Delta g d\sigma, \tag{7.14b}$$

is the approximate height anomaly, and

$$\begin{aligned} d\zeta_{dwc} &= \frac{r_P}{4\pi\gamma} \iint_{\sigma} S(\psi)[\Delta g(r_P, Q) - \Delta g_Q] d\sigma_Q \\ &\approx \frac{r_P}{4\pi\gamma} \iint_{\sigma} S(\psi) \left(\frac{\partial \Delta g}{\partial h}\right)_Q (H_P - H_Q) d\sigma_Q \end{aligned} \tag{7.14c}$$

is an additive correction for the analytical continuation.

Equation (7.14a) can be extended to the LSMSA strategy (cf. Equation 6.6a) as follows:

$$\zeta_P^{L,M} = \left(\zeta_P^{L,M}\right) + d\zeta_{dwc}^{L} + d\zeta_P^a + d\zeta_{e,P}, \tag{7.15a}$$

where:

$$\left(\zeta_P^{L,M}\right) = \frac{r_P}{4\pi\gamma} \iint_{\sigma} S^L(\psi)\Delta g d\sigma + \frac{r_P}{2\gamma} \sum_{n=2}^{M} \left(s_n^* + Q_n^L\right) \left(\frac{R}{r_P}\right)^{n+1} \Delta g_n, \tag{7.15b}$$

and

$$\begin{aligned} d\zeta_{dwc}^{L} &= \frac{r_P}{4\pi\gamma} \iint_{\sigma_0} S^L(\psi)[\Delta g(r_P, Q) - \Delta g] d\sigma_Q \\ &\approx \frac{r_P}{4\pi\gamma} \iint_{\sigma_0} S^L(\psi) \left(\frac{\partial \Delta g}{\partial h}\right)_Q [H_P - H_Q] d\sigma_Q. \end{aligned} \tag{7.15c}$$

The additive effects $d\zeta_P^a$ and $d\zeta_{e,P}$ for atmospheric and ellipsoidal effects are practically the same as for the geoid estimator. (A slightly different estimator $\hat{\zeta}_P$ was derived by Ågren 2004, Sect. 9.5.1).

The determination of the quasigeoid by analytical continuation has the advantage over Molodensky's original approach (integration over the Earth's surface, or, more precisely, the *telluroid*) that the integration surface is smooth, and there are therefore no integration problems. Also note that there is no topographic effect for the quasigeoid estimator (in contrast to the geoid estimator), implying that the quasigeoid is independent on the topographic density variations. Most European states use M. S. Molodensky's concept of normal heights for their height systems with a quasigeoid model as the reference surface, while the rest of the world relies on orthometric heights with the geoid as the zero-level. Considering the advances in data caption and theory for geoid and quasigeoid determinations, the question is which system is the best choice for the future. It is reasonable to assume that the latter concept, in contrast to the former, will always suffer from some uncertainty in the topographic density distribution, while Molodensky's approach to quasigeoid determination has a convergence problem. On the contrary, geoid and quasigeoid models computed by analytical continuation (e.g. RCR technique or KTH method) have no integration problem, and the quasigeoid can always be determined at least as accurately as the geoid. As the numerical instability of the analytical continuation is better controlled in the KTH method vs. the RCR method, we propose that any future height system be based on normal heights with a quasigeoid model computed similarly to or directly based on the KTH method (least squares modification of Stokes formula with additive corrections). However, one should also consider that the geoid, in contrast to the quasigeoid, is a level surface in the Earth's gravity potential field, which possibly is easier to interpolate between discrete computational points.

7.2 Comparison of Geoid and Quasigeoid Models

The classical formula for determining the geoid-from-quasigeoid correction (*GQC*) is (Heiskanen and Moritz 1967, Sect. 8.13)

$$GQC = N - \zeta = \frac{\bar{g} - \bar{\gamma}}{\bar{\gamma}} H_P \approx \frac{\Delta g^B}{\bar{\gamma}} H_P, \tag{7.16}$$

where N is the geoid height, ζ is the height anomaly, Δg^B is the simple Bouguer gravity anomaly, $\bar{g}$ and $\bar{\gamma}$ are the means of gravity along the plumb-line from the geoid to surface-point and normal gravity between the reference ellipsoid and normal height, respectively, and H_P is the orthometric height at the computation point P. This formula can be considerably improved in mountainous regions by adding a term with the topographic potential differences at the geoid and surface (Flury and Rummel 2009):

$$GQC \approx \frac{\Delta \bar{g}^B}{\bar{\gamma}} H_P + \frac{V_g^T - V_P^T}{\bar{\gamma}}, \tag{7.17}$$

where $\Delta \bar{g}^B$ is the refined Bouguer gravity anomaly.

Sjöberg (2010) and (2012) developed the formula further to a strict expression for a surface-gravity anomaly Δg_P^c reduced for the topographic signal with an arbitrary compensation, yielding:

$$GQC = \frac{\Delta g_P^c}{\bar{\gamma}} H_P + \frac{dV_g^T}{\gamma_0} - \frac{dV_P^T}{\gamma_Q} + res, \tag{7.18a}$$

where $dV^T = V^T - V^c$, V^c being the compensation potential, and

$$res = \int_0^{H_P} \frac{\Delta g^c}{\gamma} dh - \frac{\Delta g_P^c}{\bar{\gamma}} H_P \tag{7.18b}$$

is the residual term, which is assumed to be negligible in most situations and therefore not explicitly expressed in practical form. If the compensation model is an isostatic model, all terms in Eq. (7.18b) are small compared to those in Eq. (7.17). Equation (7.18a) is consistent with Bruns' formula and "the boundary condition", Eqs. (4.12a, 4.12b).

Today's GNSS positioning by gravity surveys make possible direct determination of the gravity disturbance. Alternatively, according to Eqs. (4.12a, 4.12b), it can be determined as a small correction to the gravity anomaly. As we will show here (see also Sjöberg 2015), the *GQC* can be determined more precisely from this quantity than from the gravity anomaly.The disturbing potential difference at the geoid and the surface point *P* can be expressed:

$$DT = T_g - T_P = T_g^{NT} - T_P^{NT} + V_g^T - V_P^T. \tag{7.19}$$

As the no-topography disturbing potential at the geoid, T_g^{NT} does not include the topographic signal, it can be expressed as a Taylor series at the surface point *P*:

$$T_g^{NT} = \sum_{k=0}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^k T^{NT}}{\partial H^k} \right)_P = T_P^{NT} - \sum_{k=1}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^B}{\partial H^{k-1}} \right)_P, \tag{7.20}$$

where we have used the relation:

$$\delta g^B = -\frac{\partial T^{NT}}{\partial H}. \tag{7.21}$$

By inserting Eq. (7.20) into Eq. (7.19) one obtains:

$$DT = -\sum_{k=1}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^B}{\partial H^{k-1}} \right)_P + V_g^T - V_P^T \tag{7.22}$$

and by considering that the *GQC* can be expressed:

$$GQC = N - \zeta = \frac{T_g}{\gamma_0} - \frac{T_P}{\gamma_Q} = \zeta \frac{\gamma_Q - \gamma_0}{\gamma_0} + \frac{DT}{\gamma_0}, \tag{7.23}$$

one finally arrives at the following strict formula for the difference:

$$GQC = \frac{\delta g_P^B}{\gamma_0} H_P + \zeta \frac{\gamma_Q - \gamma_0}{\gamma_0} + \frac{V_g^T - V_P^T}{\gamma_0} - \frac{1}{\gamma_0} \sum_{k=2}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^B}{\partial h^{k-1}} \right)_P, \tag{7.24a}$$

where:

$$\delta g^B = \delta g - A^T \tag{7.24b}$$

is the Bouguer gravity disturbance. The first two terms on the right hand-side of Eq. (7.24a) can be approximated by:

$$\frac{\delta g_P^B}{\gamma_0} H_P + \zeta \frac{\gamma_Q - \gamma_0}{\gamma_0} \approx \left(\frac{\delta g_P^B}{\gamma_0} - 2 \frac{\zeta}{r_P} \right) H_P = \Delta g_P^B \frac{H_P}{\gamma_0}, \tag{7.25}$$

so that the final equation becomes:

$$GQC \approx \Delta g_P^B \frac{H_P}{\gamma_0} + \frac{V_g^T - V_P^T}{\gamma_0} - \frac{1}{\gamma_0} \sum_{k=2}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^B}{\partial h^{k-1}} \right)_P. \tag{7.26}$$

The approximation used in Eq. (7.25) is less than 1 mm for any place on Earth, so that Eq. (7.26) can practically be regarded as a strict representation of the *GQC*.

Similar to Eq. (7.18a), the *GQC* in (7.26) can also be expressed with an arbitrary topographic compensation. The result is:

$$GQC = \frac{\delta g_P^c}{\gamma_0} H_P + \zeta \frac{\gamma_Q - \gamma_0}{\gamma_0} + \frac{dV_g^c - dV_P^c}{\gamma_0} - \frac{1}{\gamma_0} \sum_{k=2}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^c}{\partial h^{k-1}} \right)_P \tag{7.27}$$

or

$$GQC \approx \Delta g_P^c \frac{H_P}{\gamma_0} + \frac{dV_g^c - dV_P^c}{\gamma_0} - \frac{1}{\gamma_0} \sum_{k=2}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g^c}{\partial h^{k-1}} \right)_P. \tag{7.28}$$

As can be seen, the above solutions for the *GQC* require that the topographic attraction and its radial derivatives be determined. To simplify this burden, we

present below a method that includes the topographic bias. We start from the disturbing potential difference at the geoid and surface point P:

$$DT = T_g - T_P = T_g - T_P^* + T_P^* - T_P, \tag{7.29}$$

where T_P^* is the analytically continued disturbing potential at the geoid. Here the difference $T_P^* - T_g$ is the topographic bias, which for a constant topographic density distribution, was given by Eq. (5.21). Moreover,

$$T_P^* - T_P = -\sum_{k=1}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g}{\partial h^{k-1}} \right)_P, \tag{7.30}$$

so that the difference between the geoid and quasigeoid heights becomes:

$$GQC = \frac{\gamma_Q - \gamma_0}{\gamma_0} \zeta - 2\pi G\rho \left(H_P^2 + \frac{2H_P^3}{3R} \right) - \frac{1}{\gamma_0} \sum_{k=1}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g}{\partial h^{k-1}} \right)_P, \tag{7.31a}$$

or, approximately:

$$GQC \approx \frac{\Delta g_P}{\gamma_0} H_P - 2\pi G\rho \left(H_P^2 + \frac{2H_P^3}{3R} \right) - \frac{1}{\gamma_0} \sum_{k=2}^{\infty} \frac{(-H_P)^k}{k!} \left(\frac{\partial^{k-1} \delta g}{\partial h^{k-1}} \right)_P. \tag{7.31b}$$

As we can see, the only topographic information needed in this formula is the orthometric height and topographic density at the computation point. In practice, these equations can be used with only a few (if any) vertical gradients of the disturbing potential. Note that no terrain correction is needed.

7.2.1 The Geoid Versus the Quasigeoid: A Practical View

As we can see from above, the computational burden to determine the geoid and quasigeoid are pretty much the same when using Stokes' original and extended formulas, respectively. It would be more cumbersome to determine the quasigeoid by Molodensky's method, in which successive approximation is inclined to diverge for rough terrain. However, the estimation of the quasigeoid needs no topographic correction, although from a practical viewpoint, it can be attractive to smooth the observed gravity field by a remove-restore technique for the topography (but here the topographic model must not be as accurate as for geoid determination).

As a conclusion, the quasigeoid can be more precisely determined, while the geoid model always suffers from the uncertainty in the topographic model (see also Sjöberg 2013b).

On the other hand, the geoid is a geopotential surface, but the quasigeoid is not. Hence, the geoid is more suitable for interpolation between discrete

computational points, a practical fact that might be relevant for a very high resolution of the model.

7.2.2 Precise Orthometric Heights

It was argued in Chap. 3 that normal heights, in contrast to orthometric heights, can be determined without topographic mass information. See Eqs. (3.94a–c). Also, from Eq. (3.95) it is obvious that the orthometric height can be determined as a correction to the normal height by subtracting the GQC:

$$H_P = H_P^N - GQC. \tag{7.32}$$

In this way, precise orthometric heights can be determined from normal heights.

7.3 Combinations of Gravimetric and Geometric Geoid Solutions

So far we have only considered gravimetric geoid solutions. However, very significant geoid information is also provided by geometric methods. In the next section, we will first present the representation and combination of geometric geoid data, and in Sect. 7.3.2 the geometric and gravimetric types of data are combined.

7.3.1 Geometric Geoid Mapping

Geometric geoid solutions are provided, e.g. from satellite positioning, satellite altimetry and astro-geodetic levelling. Common to these methods are the derived geoidal heights that are obtained in discrete points and not directly covering a surface. Hence, there remains a mapping problem to solve the interpolation of the geoid from the point values to the surface representation. This problem can be solved in various ways, and below we will present three solutions: polynomial representation, least-squares interpolation (collocation) and Bjerhammar's interpolation technique.

(a) *Polynomial Representation:*

The planar positions in a local area are defined by the coordinates (x,y), where:

$x = R(\varphi - \varphi_0)$ and $y = (\lambda - \lambda_0)R \cos \varphi_0$. Here R is the mean Earth radius for the region, and (φ_0, λ_0) is the latitude and longitude of a central point of the region.

Let the geoid height N at the point (x,y) be represented by the polynomial

$$N(x,y) = C_0 + \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} C_{ij} x^i y^j \tag{7.33}$$

where C_0 and C_{ij} are constants to be estimated, and n_x and n_y are selected upper summation indices. If the number of geometric geoid heights available are more than $n_x n_y + 1$, the constants and their covariance matrix of Eq. (7.33) can be determined by least-squares adjustment, as explained in the following. Let us write Eq. (7.33) in the vector form

$$N(x,y) = \mathbf{h}^T \mathbf{c} \tag{7.34}$$

where the vector $\mathbf{h}$ contains the known terms of combinations $x^i y^j$, and the vector $\mathbf{c}$ includes the unknown coefficients C_0 and C_{ij}. If there are more such equations than the number of terms in $\mathbf{c}$, the system of equations

$$\mathbf{Hc} = \mathbf{N} - \boldsymbol{\varepsilon} \tag{7.35}$$

where $\boldsymbol{\varepsilon}$ is the residual vector of the matrix equation, assumed to be random with expectation zero, can be solved by adjustment by elements. The solution becomes:

$$\hat{\mathbf{c}} = \left(\mathbf{H^T H}\right)^{-1} \mathbf{H^T N} \tag{7.36a}$$

with the covariance matrix

$$\mathbf{Q}_{cc} = \sigma^2 \left(\mathbf{H^T H}\right)^{-1} \tag{7.36b}$$

where σ^2 is the variance of unit weight. Now, the solution $\hat{\mathbf{c}}$ can be used for surface mapping at any point P according to the formula

$$\widetilde{N}(x_P, y_P) = \mathbf{h_P^T} \hat{\mathbf{c}} \tag{7.37}$$

where $\mathbf{h}_P$ is the vector $\mathbf{h}$ applied to point P. The prediction variance is then given by:

$$\sigma^2_{N_P} = \mathbf{h_P^T Q_{cc} h}_P. \tag{7.38}$$

In this method various types of geometric geoid heights can be included, and if the quality of the data varies, the least squares procedure should be weighted.

The weakness of the polynomial method is that only very limited upper degrees n_x and n_y of the polynomials are possible, unless there is a dense and homogeneous surface coverage of data. If this is not the case, the surface representation is likely to oscillate unrealistically in areas not covered by observations. Consequently, the practical solution usually becomes too smooth to represent a detailed geoid

mapping. However, this method could be a starting point to represent the geoid trends in the area in combination with details estimated by the least-squares interpolation method presented next.

(b) *Least-Squares Interpolation*.

Let us assume that the possible geoid trends of the region have been determined by the polynomial method in a). For each geometric geoid observation N_i, the residual dN_i is computed by removing the trend. Then one may assume that dN_i is random with expectation zero. Now, introduce the general residual estimator or predictor at a point P:

$$\delta\widetilde{\mathrm{N}}_P = a_P^T \mathbf{dN} \tag{7.39}$$

where $\mathbf{a}_P^T$ is an arbitrary vector and $\mathbf{dN}$ is the vector of residuals dN_i. Let us denote the true residual represented by $\mathbf{d}\widetilde{\mathbf{N}}$ as ξ_P . Then the prediction error is:

$$\varepsilon_{\underset{\sim}{dN_P}} = \mathbf{a_P^T dN} - \xi_P \tag{7.40}$$

and the prediction variance becomes:

$$\sigma^2_{\underset{\sim}{dN_P}} = E\left\{\varepsilon_{\underset{\sim}{dN_P}} \varepsilon^T_{\underset{\sim}{dN_P}}\right\} = \sigma_0^2 - 2\mathbf{a_P^T q} + \mathbf{a_P^T Q a_P} \tag{7.41}$$

where

$$\sigma_0^2 = E\{\xi^2\}, \quad \mathbf{q} = E\{\mathbf{dN}\xi_P\} \quad \text{and } Q = E\{\mathbf{dNdN}^T\} \tag{7.42}$$

Differentiating Eq. (7.42) with respect to $\mathbf{a}$ and equating it to zero, the least-squares choice for $\mathbf{a}$ is obtained as:

$$\hat{\mathbf{a}}_P = \mathbf{Q^{-1}q} \tag{7.43}$$

and the least-squares prediction/interpolation variance becomes:

$$\sigma^2_{\underset{\sim}{dN_P}} = \sigma_0^2 - \mathbf{q^T}\hat{\mathbf{a}}_P \tag{7.44}$$

Hence the least-squares residual predictor is given by

$$d\hat{N}_P = \hat{\mathbf{a}}_\mathbf{P}^\mathbf{T}\mathbf{dN} = \mathbf{q^T Q^{-1} dN} \tag{7.45}$$

with the above prediction variance. This method can thus be used together with the previous polynomial trend-estimation technique. One weakness of the method is the representation of the covariance function for $\mathbf{q}$ and $\mathbf{Q}$. However, this is not a major problem for the interpolation.

(c) *Bjerhammar's Interpolation Technique*

Bjerhammar (1973, p. 324) proposes the following method for interpolation and smoothing:

$$\hat{N}_P = \frac{1}{D}\sum_{n=1}^{n_{\max}} \frac{N_i}{(d_{Pi}^2 + \kappa)^v} \tag{7.46a}$$

where

$$D = \sum_{n=1}^{n_{\max}} (d_{Pi}^2 + \kappa)^{-v} \tag{7.46b}$$

Here d_{Pi} is the distance between the points P and i, and κ is an arbitrary, small, positive smoothing constant. The constant v is a selected power of the weighting. For $\kappa = 0$, the method implies pure interpolation. This method, that for $v = 3$ resembles Poisson's integral formula, is very simple, but it does not provide an estimate of the prediction v.

7.3.2 Least Squares Combination of Gravimetric and Geometric Geoid Data

Before the gravimetric and geometric geoid estimates can be merged, it is essential to know whether there are significant biases, tilts or any other type of systematic differences between the two data sets. This question can be answered by the following test.

Let us introduce the difference $d_i = x_i - y_i$ between a gravimetric geoid height x_i and a geometric geoid height y_i. We assume that both data sets are normally distributed, and that the n number of available differences is uncorrelated with the mean value $\bar{d}$ and standard error $s_{\bar{d}}$ given by the equations

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \text{and} \quad s_{\bar{d}}^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n} (d_i - \bar{d})^2. \tag{7.47}$$

We form the test variable $T = |\bar{d}|/s_{\bar{d}}$, which is Student ($t$-) distributed with $n-1$ degrees of freedom. Hence, if $T < t_{\frac{\alpha}{2}}(n-1)$, where t is obtained from a statistical t-table at the chosen risk level α (e.g. 5%), then it is concluded that the null hypothesis (H_0: there is no bias between the data sets x_i and y_i) is accepted, otherwise H_0 is rejected.

If H_0 is rejected, one may possibly estimate the long-wavelength discrepancy between the two data sets by linear regression. For example, geoid heights determined by satellite positioning and levelling may be contaminated by systematic

errors of both the satellite system and the levelling system, but also the gravimetric geoid heights may be biased and tilted. A standard four-parameter model to account for bias and tilts between two such systems (or, more precisely for different zero- and first-degree harmonic terms) is:

$$\Delta N_i - \varepsilon_i = a + b\ \cos\varphi_i \cos\lambda_i + c\ \cos\varphi_i \sin\lambda_i + d\ \sin\varphi_i \tag{7.48a}$$

where ΔN_i is the gravimetric geoid height minus the geometric geoid height at point number i, and the coefficients a, b, c and d are the transformation parameters which can be determined together with their standard errors in a least squares adjustment from the data set ΔN_i; $i = 1, 2, \ldots, n$. Only a transformation model with all parameters being significantly different from zero should be used in the final transformation model derived from Eq. (7.48a). For a local region the simple three-parameter model

$$\Delta N_i - \varepsilon_i = a + bR(\varphi_i - \varphi_0) + (\lambda_i - \lambda_0)cR\ \cos\lambda_0 \tag{7.48b}$$

may be more suitable. Here, (φ_0, λ_0) is the pair of latitude and longitude of a selected central point of the region, and R is the mean Earth radius of the region. More generally, we can write the linear transformation as:

$$\Delta N_i - \varepsilon_i = \mathbf{a}^T\mathbf{x}, \tag{7.48c}$$

where the vector $\mathbf{a}$ contains the known transformation functions and $\mathbf{x}$ consists of the transformation parameters. For instance, in the case of five- and seven-parameter transformations, frequently:

$$\mathbf{a}^{\mathrm{T}}\mathbf{x} = x_1 + (\cos\varphi\ \cos\lambda)x_2 + (\cos\varphi\ \sin\lambda)x_3 + (\sin\varphi)x_4 + (\sin^2\varphi)x_5 \tag{7.48d}$$

and

$$\mathbf{a}^{\mathrm{T}}\mathbf{x} = (\cos\varphi\ \cos\lambda)x_1 + (\cos\varphi\ \sin\lambda)x_2 + (\sin\varphi)x_3 + \left(\frac{\sin\varphi\ \cos\varphi\ \sin\lambda}{W}\right)x_4 + \left(\frac{\sin\varphi\ \cos\varphi\ \cos\lambda}{W}\right)x_5 + \left(\frac{1 - f^2\sin^2\varphi}{W}\right)x_6 + \left(\frac{\sin^2\varphi}{W}\right)x_7, \tag{7.48e}$$

respectively, where $W = \sqrt{1 - e^2\sin^2\varphi}$ and f is the flattening of the reference ellipsoid. In general, the error component ε_i can vary from point to point, which effect can beconsidered by using appropriate weights in the least-squares adjustment.

The choice of transformation model can be judged from various tests, where the significance test of the parameters is one, and "the coefficient of determination" is another. The first test compares each set of normed least-squares estimated parameters $(\hat{x}_i/s_{\hat{x}_i})$ against the Student's t- (or tao-) distribution table to test whether the estimated x_i is significant. The second test uses the coefficient of determination

(R^2) or R-squared, which is a test statistic of the goodness-of-fit of a regression model, and it tells how well the model fits the observation points by a rational number. In case of perfect fit, the number is one, otherwise the number is smaller. Hence, the closer the number is to one, the better is the regression model. There are various expressions for R^2, all of them related to the total, and either the regression or residual sum of squares (with obvious notations):

$$SS(tot) = \sum_i (y_i - Y)^2,\ SS(reg) = \sum_i (\hat{y}_i - Y)^2 \text{ and } SS(reg) = \sum_i (\hat{y}_i - y_i)^2, \tag{7.49a}$$

respectively, where y_i and Y are an individual observation (in the present case equal to ΔN_i) and its mean value for the set of observations, respectively, and $\hat{y}_i$ is the adjusted observation. A general expression for R-squared is:

$$R^2 = 1 - \frac{SS(\text{res})}{SS(tot)}, \tag{7.49b}$$

which in linear regression equals:

$$R^2 = \frac{SS(\text{res})}{SS(tot)}. \tag{7.49c}$$

Equations (7.49b) and (7.49c) relate to the unexplained and explained fractions of the total variance, respectively. Another definition of R^2 is "the adjusted coefficient of determination", given by:

$$R_\sigma^2 = 1 - \left(1 - R^2\right)\frac{k-1}{n-k}, \tag{7.49d}$$

which better considers additional parameters in the regression than R^2 . Here k is the number of unknowns in the regression model.

Once the transformation model has been settled, it is used to correct (added to) all geometric geoid heights to remove the systematic deviations vs. the gravimetric model. The result of the transformation by one of Eqs. (7.48a–7.48e) is thus a set of approximately-independent, geometric geoid heights consistent with the gravimetric geoid heights. These discrete and corrected geometric geoid heights can now be used with any of the surface mapping methods of the previous section for geoid determination at arbitrary points. The geometric geoid height y_P, corrected for systematic errors, is finally merged with the gravimetric geoid height x_P by the weighted mean:

$$\hat{N}_P = \frac{p_x x_P + p_y y_P}{p_x + p_y} \text{ where } p_k = s_k^{-2} \text{ with } k = x, y. \tag{7.50}$$

The estimator $\hat{N}_P$ is the best solution for the geoid height according to least-squares theory. However, it does not necessarily provide the best transformation of a

geodetic height h to the orthometric or normal height. This problem (GNSS levelling) is treated in the next section.

7.3.3 GNSS Levelling

In surveying engineering, it is common to transform GNSS-derived geodetic heights h to regional/national orthometric or normal heights (H and H^N, respectively) by subtracting the geoid or quasigeoid height, respectively. However, due to the systematic differences in the triplet geoid, GNSS and levelling heights, it is more fruitful to make use of the transformation model between gravimetric, GNSS and levelling systems discussed in the previous section. However, as in this case one wants to correct the gravimetric geoid model to the geometric/GNSS model, the "corrective surface" is the negative of the transformations, e.g. provided by Eqs. (7.48a–7.48e). This problem is also related to the problem of discriminating the error sources to those related with the gravimetric geoid model and those related with the geometric geoid modelling. Moreover, in the combination of these two types of models, it is possible to improve the weight relations by employing so-called variance component estimation (e.g. Sjöberg 1983, 1984) along with the combination of the two types of models. For more details, see, e.g. Fotopolous (2003, 2013).

7.4 The Determination of W_0

7.4.1 Introduction

The level surface of the Earth's gravity field defined by the undisturbed sea level is the Gauss—Listing definition of the geoid (Gauss 1828; Listing 1873). Choosing the geoid as the global vertical datum (GVD) implies that the datum is defined by the potential (W_0) of this particular level surface of the Earth's gravity field.

Frequently, the normal potential U_1 at the selected reference ellipsoid, e.g. Geodetic Reference System 1980 (GRS80), is defined to be equal to that of the geoid, which is not well known. Then the problem is that U_1 will not be precise enough for today's need for defining W_0, as the data has considerably improved and mean sea level has been increasing by the order of 1.7 mm/year on average during the 1900s and accelerating to more than 3 mm/year today (e.g. Nicholls and Cazenave 2010). That is, although U_1 may be kept fixed to that of the GRS80 reference system, the geopotential at the geoid (W_0) frequently needs a realization that better agrees with the Gauss–Listing definition.

The advent of satellite altimetry in the 1970s provided a tool for the realization of a GVD as being the equipotential surface of the Earth's gravity field that

minimizes the sea-surface topography (SST) over all the oceans in a least-squares sense (Mather 1978). This leads to Approach I as treated in the next section, which implies a direct integration of satellite altimetry derived from sea-surface topography (SST; frequently also denoted Dynamic Ocean Topography) combined with the potential of an Earth gravitational model (EGM) over all the oceans. In contrast, Approach II consists of using the same data to first determine the size of the axes of the globally best-fitting ellipsoid to the geoid surface (called the Mean Earth Ellipsoid; MEE; Heiskanen and Moritz 1967, p. 214), followed by determining W_0 from the result. A major problem with Approach II is that satellite altimetry is only successful over the oceans, while the method requires global data. Sanchez (2012) reviews the development in the field with many references. Next we follow the presentation of Sjöberg (2013a).

In Sect. 7.4.2, Approach I is reviewed and a short discussion is provided on some of the problems; and in Sect. 7.4.3, Approach II is presented under the consideration that the zero-degree harmonic for the EGM derived geoid model is either known or unknown. This implies also that, while the above approaches assume that both the geocentric gravitational constant and the Earth's mean daily angular velocity are known (fixed), we will assume at the end that the former constant is only approximately known.

7.4.2 *Approach I: Direct Determination of* $\mathbf{W_0}$ *from Satellite Altimetry and an EGM*

7.4.2.1 Geometric and Gravimetric Geoid Heights

By satellite positioning, the geodetic height h of the Earth surface above the reference ellipsoid can be determined. Assuming also that the orthometric height (H) is known, the geometric determination of the geoid height becomes:

$$N^h = h - H. \tag{7.51}$$

For land areas, this technique for geoid-height determination is usually called GNSS/levelling, where h is determined by GNSS technology and H is the orthometric height determined by levelling and gravity. At sea, h is the geodetic height of mean sea level determined from satellite altimetry, while H is the SST, which practically is either ignored, derived by some oceanographic method or estimated from satellite altimetry and a preliminary geoid model. Importantly, for land applications, Eq. (7.51) suffers from inherited systematic errors, primarily biases, in the levelling networks, which make the formula less useful (or even useless) for solving the GVD problem (but useful for transformations from the GVD to local height systems; e.g. Sjöberg 2011). For ocean areas, Eq. (7.51) is most important

despite the fact that the SST is frequently unknown and simply neglected. This is because, except for the long-wavelength gravity-field features as determined by satellite data, there is very sparse gravity-related data available from other sources at sea.

As an alternative, the geoid height can also be estimated *gravimetrically* from an EGM, such as EGM2008 (Pavlis et al. 2012). Neglecting the atmosphere, the Earth's gravity potential outside the topographic masses can be represented as an external type series of spherical harmonics:

$$\widetilde{W}(r,\theta,\lambda) = \frac{GM_1}{r}\left[1 + \sum_{n=2}^{n_{\max}} \sum_{m=-n}^{n} A_{nm}\left(\frac{R}{r}\right)^{n+1} Y_{nm}(\theta,\lambda)\right] + \Phi(r,\theta,\lambda), \quad (7.52)$$

where (r,θ,λ) are the geocentric radius, co-latitude and longitude of the computational point, GM_1 is an adopted value for the geocentric gravitational constant, Φ is Earth's rotational potential, each spectral potential component A_{nm} is determined from a global set of gravity related data by harmonic analysis up to the chosen degree and order $n_{\max}$ at the reference radius R and $Y_{nm}(\theta,\lambda)$ is the surface spherical harmonic of degree n and order m. One notices that there are no first-degree terms in Eq. (7.52), which implies that the origin of the coordinate system is selected at the Earth's gravity centre. In a similar way, a normal gravity-field potential can also be expressed as a harmonic series of a gravitational potential plus the rotation potential:

$$U(r,\theta,\lambda) = \frac{GM_1}{r}\left[1 + \sum_{n=2}^{n_{\max}} \left(\frac{R}{r}\right)^{n} \sum_{m=-n}^{n} B_{nm} Y_{nm}(\theta,\lambda)\right] + \Phi(r,\theta,\lambda), \quad (7.53)$$

whose potential is constant (say, U_1) on a chosen (level) reference ellipsoid with mass M_1. Then one obtains the following series for *the disturbing potential*:

$$T(r,\theta,\lambda) = \widetilde{W}(r,\theta,\lambda) - U(r,\theta,\lambda) = \frac{GM_1}{r}\sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n} \sum_{m=-n}^{\infty} C_{nm} Y_{nm}(\theta,\lambda), \quad (7.54)$$

where $C_{nm} = A_{nm} - B_{nm}$. In practice, one tries to choose the constant GM_1 as the best available estimate of the geocentric gravitational constant. From now on, the disturbing potential estimate from the EGM in Eq. (7.54), limited to some maximum degree, will be denoted by T^{EGM}.

A direct way to estimate the geoid potential (W_0) is to apply Eq. (7.52) at the radius vector r_g of the geoid (e.g. Dayoub et al. 2012):

$$\hat{W}_g = \hat{W}_0 = \hat{W}(r_g,\theta,\lambda), \quad (7.55)$$

where

$$r_g = r_g(\theta, \lambda) = r_1(\theta) + N(\theta, \lambda). \tag{7.56}$$

Here, $r_1(\theta)$ is the geocentric radius vector of the defined reference ellipsoid, and $N(\theta, \lambda)$ is the related geoid height (which we frequently will abbreviate with N), which can be estimated geometrically from satellite altimetry in ocean areas according to Eq. (7.51).

Alternatively, one may start by applying Bruns' formula for the normal potential at the geoid (U_g; Heiskanen and Moritz 1967, p. 84):

$$T_g = W_0 - U_g = W_0 - (U_1 - \gamma_1 N), \tag{7.57}$$

which leads to (see also Sacerdote and Sanzó 2004):

$$W_0 = U_1 - \gamma_1 N + T_g, \tag{7.58}$$

where U_1 (= constant) and γ_1 are the normal potential and gravity on the reference ellipsoid, respectively. [Actually, $\gamma_1 = \gamma_1(\theta)$, but we will frequently just use the short notation. In practice, $T^{EGM} = T^{EGM}(r_1(\theta), \theta, \lambda)$ can be used for representing the potential T_g, i.e. the approximation error of using $r_1(\theta)$ for r_g is usually negligible.] From Eq. (7.58), it thus follows that the geoid height above the reference ellipsoid is given by:

$$N = \frac{T^{EGM}}{\gamma_1} - \frac{\Delta W_0}{\gamma_1}, \tag{7.59}$$

where $\Delta W_0 = W_0 - U_1$. Equation (7.59) shows that in general the geoid height

$$N^{EGM} = \frac{T^{EGM}}{\gamma_1}, \tag{7.60}$$

determined by the EGM, lacks the unknown correction $-\Delta W_0/\gamma_1$, which must be determined from geometric data (e.g. by satellite altimetry at sea; see Sect. 2.2). For a detailed discussion of the determination of the absolute geoid height from an EGM, see Smith (1998).

It is important to remember that, on the continents, T_g is the disturbing potential inside the topographic masses, and its computation therefore needs a correction for the topography. That is, the harmonic series for the geoid height in Eq. (7.60) needs a correction for the analytical downward continuation error or topographic bias of T_g^{EGM} (Sjöberg 1977 and 2007; Martinec 1998, Sects. 7.3–7.4; Ågren 2004), which is −5 cm for the zero-degree harmonic (Sjöberg 2001).

7.4.2.2 Direct Estimates of W_0

Equations (7.55) and (7.58) are the bases for the direct determination of the geoid potential. One estimator of W_0 is obtained by averaging $\hat{W}_g$ of Eq. (7.55) over the oceans. In two other estimates, one may take advantage of the geodetic height determined by satellite altimetry by Eq. (7.51) and the disturbing potential (determined by an EGM) at the surface point on the undisturbed sea level (assumed to be the geoid surface), possibly corrected for SST, yielding the following result for a first order Taylor expansion (e.g. Sacerdote and Sansó 2004):

$$W_0 = U_g + T_g = U_1 - \gamma h_g + T_g^{EGM}, \tag{7.61}$$

where $h_g = h - SST = N^h$ is the geoid height determined from satellite altimetry. By taking the mean value over the ocean area (σ_1) of such point-wise estimates for W_0, one obtains the following estimators of the geopotential at the geoid:

(1) Averaged geopotential on the ocean

$$\hat{W}_0 = U_1 + \frac{1}{\sigma_1} \iint_{\sigma_1} \left(T_g^{EGM} - \gamma_1 h_g \right) d\sigma \tag{7.62}$$

and

(2) Minimized SST (Sacerdote and Sansó 2004)

$$\hat{W}_0 = U_1 + \frac{\iint_{\sigma_1} \left[\frac{T_P^{EGM} - \gamma h_g}{\bar{\gamma}_P^2} \right] d\sigma}{\iint_{\sigma_1} \left[\frac{1}{\bar{\gamma}_P^2} \right] d\sigma}, \tag{7.63}$$

where subscript P refers to surface point P, and $\bar{\gamma}_P$ is the mean normal gravity along the normal height at P.

In a similar way, the estimator of Eq. (7.55) can be averaged over the oceans.

7.4.2.3 Discussion

If the integration area in Eq. (7.62) were the whole sphere, the disturbing potential determined by the EGM would vanish, and the estimator would be U_1 minus the global average of the geometric geoid height (times γ). In all other cases, the solution depends on the EGM, which includes both commission and omission errors. For instance, when σ_1 is the area of the oceans, one can expect that the RMS-geoid error of EGM2008 complete to degree 2159 is about 5–6 cm (Pavlis et al. 2012), corresponding to an uncertainty in W_0 of about 0.5–0.6 $(\mathrm{m/s})^2$, and this value is in agreement with current accuracy in determining W_0 (Sanchez 2012).

7.4.3 *Approach II: Joint Determination of* W_0 *and the MEE Parameters*

7.4.3.1 Introduction

Here the geometric approach to determine W_0 will be presented under the assumption that the mean angular velocity of the Earth's daily rotation (ω) is known, and the problem is to estimate both the dimensions [i.e. semi-major and -minor axes a and b (or eccentricity e) of the globally best-fitting ellipsoid (= the Mean Earth Ellipsoid, MEE)], as well as the geoid potential W_0 in a joint adjustment from the available geoid surface estimates derived from satellite altimetry and an EGM. The ideal normal potential U_0 of the MEE is given by four parameters, namely, *a, b, GM* and ω (see Sect. 3.2.2 or Heiskanen and Moritz 1967, Eq. 2.61):

$$U_0 = \frac{GM}{\sqrt{a^2 - b^2}} \arctan \frac{\sqrt{a^2 - b^2}}{b} + \frac{\omega^2 a^2}{3}, \tag{7.64}$$

or, by using the substitution $\arctan\left(e/\sqrt{1-e^2}\right) = \arcsin(e)$,one obtains:

$$U_0 = \frac{GM}{ae} \arcsin(e) + \frac{\omega^2 a^2}{3}, \tag{7.65}$$

and this potential is also the best choice for W_0. As will be shown, the geocentric gravitational constant *GM* is not part of the adjustment, (but it could be indirectly estimated from the adjustment results). The general computational procedure is outlined below.

Similar to Sect. 7.4.2, the first the normal potential U_1 of a preliminary reference ellipsoid with geometric parameters a_1 and $b_1 = a_1\sqrt{1-e_1^2}$ and geocentric gravitational constant GM_1 is given:

$$U_1 = \frac{GM}{a_1 e_1} \arcsin(e_1) + \frac{\omega^2 a_1^2}{3}, \tag{7.66}$$

Where all parameters are chosen. The radius vector of the surface of this reference ellipsoid is given by:

$$r_1(\beta) = a_1\sqrt{1 - e_1^2 \sin^2 \beta}, \tag{7.67}$$

where β is the reduced latitude.

According to Heiskanen and Moritz (1967, p. 214), the globally best fitting reference ellipsoid, i.e. the Mean Earth ellipsoid (MEE), is the ellipsoid, whose mass is the same as that of the real Earth (requires that M_1 is assumed to be the

Earth's mass), and the axes are such that the global mean square of the geoid height (N) is a minimum:

$$\frac{1}{4\pi}\iint_{\sigma} N^2 d\sigma = \min.(a, e), \tag{7.68}$$

where σ is the unit sphere. The best-known value for GM has a standard error on the order of 0.8 $m^{-3}s^{-2}$ (Groten 2004), which corresponds to uncertainties in W_0 and the geoid height of 0.1 m^2s^{-2} and 1 cm, respectively. As the present uncertainty in W_0 is about 0.5 m^2s^{-2} (e.g. Sanchez 2012), the fixing of GM to the best-known value may improve W_0 by adjusting just for the ellipsoidal geometry parameters by the following geometric approach. This implies that, once a and e have been fixed, U_0 (and yielding $W_0 = U_0$) follows from Eq. (7.65).

However, to be more precise in the approach to follow, it is not the integral in Eq. (7.68) that is to be minimized, but it is the mean square discrepancy between the radius vector of the geoid surface estimated by $r_1(\beta) + N$, where $r_1(\beta)$ is the radius vector of the reference ellipsoid related to the geoid estimate N, and the radius vector $r_E(a, e, \beta)$ of a general reference ellipsoid that should be optimized. Hence, mathematically the problem could be expressed as:

$$J = \frac{1}{4\pi}\iint_{\sigma} [r_1(\beta) + N - r_E(a, e, \beta)]^2 d\sigma = \min.(a, e), \tag{7.69a}$$

where:

$$r_E(\beta) = a\sqrt{1 - e^2 \sin^2 \beta}. \tag{7.69b}$$

Once the ellipsoidal parameters a and e of the MEE have been fixed by solving Eq. (7.69a), the normal potential at the MEE, i.e. U_0 of Eq. (7.65), can be computed, provided that GM is (sufficiently well) known, and this value should also be the estimate for the geopotential value at the geoid, i.e.

$$W_0 = U_0. \tag{7.70}$$

However, one problem with this approach is that the present-day uncertainty in GM contributes to about 20% of the uncertainty in W_0 (see Groten 2004). In addition, the main problem to optimize the target function J is that the absolute geoid height is not well-known globally, but there are only relative geoid models, such as those expressed by an EGM. However, as stated above, satellite altimetry can provide a geometric estimate of the absolute geoid height over the oceans under the assumption that the SST is known with sufficient accuracy, but such an incomplete integration area for J in Eq. (7.69a) would only lead to the best-fitting reference ellipsoid and W_0 estimated for the ocean areas. Dayoub et al. (2012) applied the above technique, and they compared the preliminary geoid surfaces

given by EGM2008 and a satellite altimetry model in coastal areas and concluded that the two surfaces agree well (without specifying the magnitude of the agreement), and they directly filled-in the geoid heights for the land areas by EGM2008 geoid heights. However, the EGM-derived geoid height needs the correction $-\Delta W_0/\gamma_1$ of Eq. (7.59), which must be estimated. If the correction is fixed to a preliminary value (which apparently was the case in Dayoub 2012), it means that also $W_0 = U_1 + \Delta W_0$ has been (more or less) fixed. However, to avoid fixing W_0 to an *a priori* value, the problem can be solved by augmenting the target function of Eq. (7.69a) by the unknown parameter ΔW_0. This implies that *a, e* and ΔW_0 are determined in a combined adjustment.

7.4.3.2 The Combined Adjustment Approach

In the previous section, the target function J was based on the assumptions that the geocentric gravitational constant is known and agrees with that of the normal potential U_1, and the estimated geoid surface is continuous and known all over the Earth. In reality, we have seen that neither of these assumptions is correct. From satellite altimetry, the geoid height is known only over the oceans, and the EGM-geoid height lacks the term $\Delta W_0/\gamma_1$ as presented in Eq. (7.59). In the approach that follows, we are not primarily concerned with the unknown *GM*, but it is sufficient to consider the extra unknown $x = -\Delta W_0 = U_1 - W_0$. We will assume that the reference ellipsoids for N^h and N^{EGM} are the same (with geometric parameters a_1 and e_1), and the radius vector is given by Eq. (7.67). Then the augmented target function reads:

$$I = pI_1(a,e) + (1-p)I_2(x,a,e) + I_3(x,a,e) = \min(x,a,e), \tag{7.71a}$$

where

$$I_1(a,e) = \iint_{\sigma_1} \left[N^h + r_1(\beta) - r_E(a,e,\beta)\right]^2 d\sigma \tag{7.71b}$$

$$I_2(x,a,e) = \iint_{\sigma_1} \left[x/\gamma_1 + N^{EGM} + r_1(\beta) - r_E(a,e,\beta)\right]^2 d\sigma \tag{7.71c}$$

and

$$I_3(x,a,e) = \iint_{\sigma_2} \left[x/\gamma_1 + N^{EGM} + r_1(\beta) - r_E(a,e,\beta)\right]^2 d\sigma. \tag{7.71d}$$

Here σ_1 and σ_2 are those parts of the unit sphere that are covered by ocean and land, respectively, $x = -\Delta W_0 = U_1 - W_0$ (see Eq. (7.59)) and $r_E(a,e,\beta)$, given by Eq. (7.69b), is the radius vector of the MEE, whose parameters a and e are

unknowns, and p is a fixed number in the range $0 \leq p \leq 1$ that weights the contributions from N^h and N^{EGM} over the ocean areas.

If p is set to 0, implying that only EGM data are used, the solution is singular. This is obvious, as in this case one tries to solve the problem with only relative geoid heights given by the EGM. Alternatively, if $p = 1$ (i.e. only the satellite-altimetry -derived geoid height is employed over the oceans, while the EGM is utilized only over land), the solution discards the information from the EGM over the oceans. On the contrary, below we suggest that in the application of Eqs. (7.71a–7.71d), the choice of p should be based on the *a priori* variances κ_1^2 and κ_2^2 of the satellite altimetry and EGM-derived geoid heights, respectively, yielding $p = \kappa_2^2/(\kappa_1^2 + \kappa_2^2)$.

The least-squares condition for the unknowns x, a and e, as specified by Eq. (7.71a), is satisfied by the three equations

$$\frac{\partial I}{\partial x} = 0, \ \frac{\partial I}{\partial a} = 0 \text{ and } \frac{\partial I}{\partial e^2} = 0, \tag{7.72}$$

from which the unknowns can be determined, provided that the equations are independent.If $\hat{x}$ is the solution for x, and U_1 is the *a priori* value for the ellipsoidal normal potential, the geoid potential estimate finally follows from:

$$\hat{W}_0 = U_1 - \hat{x}, \tag{7.73}$$

And, by re-inserting the estimates for $W_0 = U_0$, a and e, into Eq. (7.65), a new estimate for GM becomes:

$$G\hat{M} = (\hat{W}_0 - \hat{a}^2\omega^2/3)\frac{\hat{a}\hat{e}}{\arcsin(\hat{e})} \approx (\hat{W}_0 - \hat{a}^2\omega^2/3)\hat{a}\left(1 - \frac{\hat{e}^2}{6}\right), \tag{7.74}$$

But this estimate is probably poor compared to estimates based on satellite-laser ranging, etc. (e.g. Ries et al. 1992).

This concludes the principle of the combined approach.

The assumptions, data, the problem and its solution can be summarized as follows:

Given (fixed parameters): ω and a reference/level ellipsoid with dimensions a_1 and $b_1 = a_1\sqrt{1 - e_1^2}$, as well as a normal gravity field with values $U_1 =$ constant and normal gravity, $\gamma_1 = \gamma_1(\beta)$ on the surface of the ellipsoid.

Observations: N^{EGM} (globally) and N^h (on the oceans; σ_1); both types of geoid heights refer to the defined reference ellipsoid.

Problem: Minimize the global mean square difference between radius vectors of the surfaces of the geoid and an arbitrary reference ellipsoid (with parameters a and e) by varying these parameters and the additional unknown $x(= U_1 - W_0)$ until the minimum is reached. As there are two types of differences available in the region σ_1, these differences must be weighted in one way or another. The mathematical

formulation of the problem is given by the target function described in Eqs. (7.71a–7.71d).

Solution: The solution follows from Eqs. (7.72).

Details and hints on the practical solution to the latter approach are given in Sjöberg (2013a, b).

7.4.4 Final Remarks

The geoid potential W_0 can be directly estimated as a correction to the normal potential of the reference ellipsoid and the oceanic average of the difference between geoid heights from satellite altimetry and an EGM. Systematic errors in the data propagate into biased solutions, which are therefore sensitive to the chosen data. Typical systematic error sources are lacking SST information for the satellite altimetry (which locally may reach several decimetres to a metre) and truncation error in the EGM-derived geoid height. The effects of these error types will change with the area of integration. The latter error source would vanish, if the averaging is extended to the whole Earth, but that would require geometric geoid heights in continental regions, which are prone to other types of biases.

This technique cannot be used to estimate the axes of the MEE, but their relationships can be conditioned by Eq. (7.66), once $U_0 = W_0$ has been fixed.

The alternative technique to solve for W_0 is to first determine the axes of the MEE and then find the geoid potential. The main problem with this technique is that it propagates a bias from the unknown zero-degree harmonic of the gravimetric geoid height into the solution. This problem is solved by the new technique discussed in Sect. 7.4.3, where W_0 (or its correction from the potential of the preliminary reference ellipsoid) is included as an extra unknown. Again, a main problem is due to the more-or-less unknown SST corrections in the ocean areas, which may be of the order of some metres.

7.5 Spectral Smoothing and Combination

7.5.1 Introduction

In this section, our first goal is to downward continue satellite derived gravity gradients (SGG data) or airborne-gravity disturbances to sea level by spectral smoothing, and, second, to combine such data, e.g. with terrestrial gravity data and or with an EGM. The goal is also to estimate the geoid height from such data. The estimates are determined in an optimum way. The tool to do so is the spectral combination technique, which method has the following advantages to previous methods for incorporating SGG or air-gravimetry data in gravity and geoid

determination: (1) it is based on a least squares combination of all data in the spectral domain, which yields a minimum expected mean square error of the estimated quantity (which satisfies the criterion on optimization), and (2) the well-known instability in solving an integral equation for inverting SGG data is replaced by pure forward integration with smoothing for the observation noise. Hence, spectral combination is an integral formula with spectral weighting of the observations, where the weights are determined by least squares. Some typical cases of spectral combination of SGG data with terrestrial gravity data and an EGM will be numerically studied by simulation, the major properties of the resulting integral kernels will be analysed, and the global mean square error (MSE) of each method of combination will also be derived.

7.5.2 *Spectral Smoothing of SGG Data*

Here the problem of recovering/estimating the gravity disturbance and geoid height from the second-order radial derivative of the disturbing potential by spectral smoothing is investigated. The estimators that continue the SGG data down from space to sea level, called spectral smoothing, are presented both in the spectral domain as infinite series and, more practically, as integral formulas.

7.5.2.1 Gravity Disturbances

Let us assume that the vertical component of the SGG data, T_{rr}, is available on a sphere of radius $r > R =$ sea level radius. (In practice these data are not primarily homogeneously distributed on a sphere, but some method, like least- squares prediction/LSC; Heiskanen and Moritz 1967, pp. 268–270, should be used to arrange the data in this form.) Also assumed known are the signal-degree variances at sea level of the gravity disturbance ($c_{\delta g,n}$), where index n denotes the spectral degree. The task is to determine the gravity disturbance at sea level from these data in an optimum way (e.g. the application of LSC; e.g. Migliaccio et al. 2010; Pail et al. 2010; Schuh et al. 2010; Bruinsma et al. 2010; Metzler and Pail 2005). Using the notations $T_{rr,n}$, δg_n and $\varepsilon_{rr,n}$ for the (Laplace harmonic) spectra of the observed second-order radial derivative of the disturbing potential (at level r), the error-free gravity disturbance (at sea level) and the observation error, respectively, a general gravity-disturbance estimator at sea level can be expressed in the spectral form by the series:

$$\delta\tilde{g} = \sum_{n=0}^{\infty} A_n k_n T_{rr,n}, \tag{7.75a}$$

with the error

$$\varepsilon_{\delta g} = \sum_{n=0}^{\infty} A_n k_n \varepsilon_{rr,n} \tag{7.75b}$$

and bias

$$b = E\{\delta\tilde{g}\} - \delta g = \sum_{n=0}^{\infty} (k_n \kappa_n - 1)\delta g_n. \tag{7.75c}$$

where $A_n = R/(n+2)$ and k_n are spectral weights, to be selected such that the expected mean square error of Eq. (7.76) is a minimum. In the last equation, we have applied the equivalence $A_n \mathrm{E}\{T_{rr,n}\} = \kappa_n \delta g_n$, where $\kappa_n = (R/r)^{n+3}$, $\mathrm{E}\{\}$ is the statistical expectation operator, and we assume that $\varepsilon_{rr,n}$ has the expectation zero and degree variance $\sigma^2_{rr,n}$. Then it follows that the expected global mean-square error of $\delta\tilde{g}$ (being the sum of global averages of variance and bias squared) becomes:

$$m^2_{\delta g} = \mathrm{E}\left\{\frac{1}{4\pi}\iint_{\sigma} \varepsilon^2_{\delta g} d\sigma\right\} = \sum_{n=0}^{\infty}\left[A_n^2 k_n^2 \sigma^2_{rr,n} + (k_n\kappa_n - 1)^2 c^2_{\delta g,n}\right] \tag{7.76}$$

where $c^2_{\delta g,n}$ is the single degree variance of the gravity disturbance and the first and second terms on the right hand-side are the averaged variance and bias squared, respectively. (Note that all possible degree-correlations are averaged out when taking the global mean; see Appendix.) By differentiating the mean-square error with respect to k_n and equating it to zero, one obtains the least-squares solution for the weights:

$$\hat{k}_n = \frac{\kappa_n c^2_{\delta g,n}}{D_n}, \text{ where } D_n = \kappa_n^2 c^2_{\delta g,n} + A_n^2 \sigma^2_{rr,n}, \tag{7.77}$$

and the optimum gravity-disturbance estimator at sea level becomes:

$$\delta\hat{g} = \sum_{n=0}^{\infty} A_n \frac{c^2_{\delta g,n}\kappa_n}{D_n} T_{rr,n}, \tag{7.78a}$$

or

$$\delta\hat{g} = \sum_{n=0}^{\infty} f_n A_n \frac{T_{rr,n}}{\kappa_n}; \quad f_n = \left[1 + A_n^2 \sigma^2_{rr,n}/(\kappa_n^2 c^2_{\delta g,n})\right]^{-1}, \tag{7.78b}$$

where f_n is a degree-smoothing factor, which damps the directly downward-continued signal and error of the gradiometry observations.

The expected MSE becomes:

$$m_{\delta\hat{g}}^2 = \sum_{n=0}^{\infty} A_n^2 \frac{\sigma_{rr,n}^2 c_{\delta g,n}^2 \kappa_n^2}{D_n}. \tag{7.78c}$$

or

$$m_{\delta\hat{g}}^2 = \sum_{n=0}^{\infty} \kappa_n^2 c_{\delta g,n}^2 \left(1 - \frac{\kappa_n^2 c_{\delta g,n}^2}{D_n}\right) < \sum_{n=0}^{\infty} \kappa_n^2 c_{\delta g,n}^2. \tag{7.78d}$$

Equations (7.78c and 7.78d) show that the MSE is always finite and approaches zero with the gradiometry data error. As the observed satellite gradiometry data are given in the spatial domain, a more practical form of the estimator is given by the integral formula:

$$\delta\hat{g} \approx \frac{R}{4\pi} \iint_{\sigma} H(\kappa, \psi) T_{rr} d\sigma, \tag{7.79a}$$

where the approximation stems from the truncation of the kernel function to a finite degree n_0 when expressing it in a series of Legendre polynomials, $P_n(\cos\psi)$:

$$H(\kappa, \psi) = \sum_{n=0}^{\infty} \frac{2n+1}{n+2} \frac{c_{\delta g,n}^2 \kappa_n}{D_n} P_n(\cos\psi). \tag{7.79b}$$

[Readers may convince themselves that Eq. (7.78a) is the spectral form of Eq. (7.79a) by expressing T_{rr} as a harmonic series, inserting Eq. (7.79b) for $H(\kappa, \psi)$ and considering the orthogonality of the harmonics when integrated over the sphere.]

Alternatively, by substituting the kernel function $H(\kappa, \psi)$ by:

$$K(\kappa, \psi) = \sum_{n=0}^{\infty} \frac{(2n+1)}{(n+2)} \frac{(n-1)}{(n+1)} \frac{c_{\delta g,n}^2 \kappa_n}{D_n} P_n(\cos\psi), \tag{7.80}$$

Equation (7.79a) is substituted by the gravity anomaly estimator

$$\Delta\hat{g} = \frac{1}{4\pi} \iint_{\sigma} K(\kappa, \psi) T_{rr} d\sigma, \tag{7.81}$$

with the expected MSE

$$m_{\Delta\hat{g}}^2 = \sum_{n=0}^{\infty} \left[\frac{R(n-1)}{(n+2)(n+1)}\right]^2 \frac{\sigma_{rr,n}^2 c_{\Delta g,n}^2 \kappa_n^2}{D_n}. \tag{7.82}$$

7.5.2.2 Geoid Heights

The presented theory can be developed further so that the SGG data are used to directly estimate the geoid. For this purpose we introduce the geoid height in the spectral domain (N_n), which is related to the gravity-disturbance spectrum by the formula (Heiskanen and Moritz 1967; Eqs. 2.153 and 2.144)

$$N_n = \frac{1}{\gamma_0}\frac{R}{n+1}\delta g_n\,; \quad n = 0, 1, 2, \ldots, \tag{7.83}$$

where γ_0 is normal gravity at the reference ellipsoid. Then, it follows from Eq. (7.78a) that the optimum geoid height estimator is given by:

$$\hat{N} = \frac{R}{\gamma_0}\sum_{n=0}^{\infty}\frac{A_n\hat{k}_n}{n+1}T_{rr,n}, \tag{7.84a}$$

or, expressed in the space domain:

$$\hat{N} = \frac{R^2}{4\pi\gamma_0}\iint_{\sigma} E(\kappa,\psi)T_{rr}d\sigma, \tag{7.84b}$$

where the spectral weights $\hat{k}_n$ were given in Eq. (7.77), and

$$E(\kappa,\psi) = \sum_{n=0}^{\infty}\frac{2n+1}{(n+1)(n+2)}\frac{c^2_{\delta g,n}\kappa_n}{D_n}P_n(\cos\psi). \tag{7.84c}$$

The expected MSE of $\hat{N}$ becomes:

$$m^2_{\hat{N}} = \left(\frac{R}{\gamma_0}\right)^2\sum_{n=0}^{\infty}\frac{A_n^2\sigma^2_{rr,n}c^2_{\delta g,n}\kappa_n^2}{(n+1)^2D_n}. \tag{7.85}$$

7.5.3 *Spectral Combination of Satellite Gravity-Gradiometry Data and an Earth Gravitational Model*

The spectral combination of SGG data with an EGM will be treated as solutions with (Sect. 7.5.3.1) and partly without (Sect. 7.5.3.2) the extra information given by the signal degree variances (c_n).

7.5.3.1 Spectral Combination Using All Signal Degree Variances

Consider the general geoid height estimator estimated by the SGG observable $T_{rr,n}$, on the sphere of radius $r > R$ = sea level radius, and the corresponding observable provided by an EGM, T_n, applied at sea level:

$$\tilde{N} = \frac{1}{\gamma_0}\sum_{n=2}^{\infty} B_n k_n T_{rr,n} + \frac{1}{\gamma_0}\sum_{n=2}^{M} a_n T_n, \tag{7.86}$$

where $B_n = R/[(n+1)(n+2)]$, k_n and a_n are arbitrary spectral weights, and M is the maximum degree of the EGM. The first sum in the estimator is the contribution from the SGG data, and the second sum is the contribution from the EGM. Introducing the random errors $\varepsilon_{rr,n}$ and ε_n for $T_{rr,n}$ and T_n, respectively, the error of $\tilde{N}$ follows from Eqs. (7.86) and (7.83):

$$\varepsilon_{\tilde{N}} = \tilde{N} - N = \frac{1}{\gamma_0}\sum_{n=2}^{M}\left\{B_n k_n \varepsilon_{rr,n} + a_n \varepsilon_n + (\kappa_n k_n + a_n - 1)T_n\right\} + \frac{1}{\gamma_0}\sum_{n=M+1}^{\infty}\left[B_n k_n T_{rr,n} + B_n k_n \varepsilon_{rr,n} - T_n\right] \tag{7.87}$$

Again, as in the Sect. 7.5.2.1, the factor κ_n stems from converting the observation signal and error spectra from the observation radius r to computation radius R. Assuming that $\varepsilon_{rr,n}$ and ε_n are uncorrelated with expectations zero and expected degree variances $\sigma^2_{rr,n}$ and $d\tau_n$, respectively, one obtains the following expected global MSE of $\tilde{N}$:

$$m^2_{\tilde{N}} = \frac{1}{\gamma_0^2}\sum_{n=2}^{M}\left[B_n^2 k_n^2 \sigma^2_{rr,n} + a_n^2 d\tau_n^2 + (\kappa_n k_n + a_n - 1)^2 \tau_n^2\right] + \frac{1}{\gamma_0^2}\sum_{n=M+1}^{\infty}\left[B_n^2 k_n^2 \sigma^2_{rr,n} + (\kappa_n k_n - 1)^2 \tau_n^2\right], \tag{7.88}$$

where τ_n^2 is the signal-degree variance of T_n. Equation (7.87) shows that the general estimator $\tilde{N}$ is biased for all degrees (i.e. for each degree there is a finite contribution to the geoid error even for perfect data). Differentiating Eq. (7.88) with respect to k_n and a_n, and equating each of these equations to zero, one obtains two equations for each degree $n \le M$:

$$k_n\left(B_n^2 \sigma^2_{rr,n} + \kappa_n^2 \tau_n^2\right) + a_n \kappa_n \tau_n^2 = \kappa_n \tau_n^2 \tag{7.89a}$$

and

$$k_n \kappa_n \tau_n^2 + a_n\left(\tau_n^2 + d\tau_n^2\right) = \tau_n^2, \tag{7.89b}$$

yielding the least squares solutions for the spectral weights:

$$\hat{k}_n = \frac{\kappa_n \tau_n^2 d\tau_n^2}{F_n} \quad \text{and} \quad \hat{a}_n = \frac{B_n^2 \sigma_{rr,n}^2 \tau_n^2}{F_n}, \tag{7.90a}$$

where:

$$F_n = \kappa_n^2 \tau_n^2 d\tau_n^2 + B_n^2 \sigma_{rr,n}^2 \left(\tau_n^2 + d\tau_n^2\right). \tag{7.90b}$$

For $n > M$, one obtains in a similar way by differentiating Eq. (7.88) with respect to k_n and equating to zero:

$$\hat{k}_n = \frac{\kappa_n \tau_n^2}{F_n} \text{ with } F_n = B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 \tau_n^2. \tag{7.91}$$

Hence, Eqs. (7.86), (7.90a), (7.90b) and (7.91) lead to the following least squares solution for the geoid height:

$$\hat{N} = \frac{1}{\gamma_0} \sum_{n=2}^{n_0} B_n \hat{k}_n T_{rr,n} + \frac{1}{\gamma_0} \sum_{n=2}^{M} \hat{a}_n T_n, \tag{7.92}$$

where $|\hat{a}_n| < 1$, implying a damping of the EGM signal. The MSE becomes:

$$m_{\hat{N}}^2 = \frac{1}{\gamma_0^2} \sum_{n=2}^{M} \frac{B_n^2 \sigma_{rr,n}^2 \tau_n^2 d\tau_n^2}{F_n} + \frac{1}{\gamma_0^2} \sum_{n=M+1}^{n_0} \frac{B_n^2 \sigma_{rr,n}^2 \tau_n^2}{F_n}. \tag{7.93}$$

For a regional application, Eq. (7.92) can be re-written as the sum of an integral formula for the SGG data and a harmonic series of the EGM:

$$\hat{N} = \frac{R^2}{4\pi\gamma} \iint_{\sigma} F(\kappa, \psi) T_{rr} d\sigma + \sum_{n=2}^{M} \frac{B_n \sigma_{rr,n}^2 \tau_n^2}{F_n} T_n, \tag{7.94a}$$

where:

$$F(\kappa, \psi) = \sum_{n=2}^{r_0} \frac{2n+1}{(n+1)(n+2)} \hat{k}_n P_n(\cos\psi). \tag{7.94b}$$

7.5.3.2 Spectral Combination Unbiased through Degree M

In practice, we should not expect the degree variances τ_n^2 to be known, but they must be estimated. However, they can be eliminated from the estimator, at least to

degree M. Also in this case, we start from the general estimator of Eq. (7.86) with the global mean square error given by Eq. (7.88). By setting

$$k_n \kappa_n + a_n - 1 = 0 \text{ for } 2 \le n \le M, \tag{7.95}$$

the dependence on τ_n^2 will be eliminated for these degrees. The general geoid estimator and its mean square error thus reads:

$$\tilde{N} = \frac{1}{\gamma} \sum_{n=2}^{M} \left[B_n k_n T_{rr,n} + (1 - k_n \kappa_n) T_n \right] + \frac{1}{\gamma} \sum_{n=M+1}^{\infty} B_n k_n T_{rr,n} \tag{7.96a}$$

and

$$m_{\tilde{N}}^2 = \frac{1}{\gamma^2} \sum_{n=2}^{M} \left[B_n^2 k_n^2 \sigma_{rr,n}^2 + (1 - k_n \kappa_n)^2 d\tau_n^2 \right] + \frac{1}{\gamma^2} \sum_{n=M+1}^{\infty} \left[(k_n \kappa_n - 1)^2 \tau_n^2 + B_n^2 k_n^2 \sigma_{rr,n}^2 \right]. \tag{7.96b}$$

Differentiating Eq. (7.96b) with respect to k_n and equating to zero leads to the least squares solution

$$\hat{k}_n = \frac{\kappa_n d\tau_n^2}{B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 d\tau_n^2} \text{ for } 2 \le n \le M \tag{7.97a}$$

and

$$\hat{k}_n = \frac{\kappa_n \tau_n^2}{B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 \tau_n^2} \text{ for } n > M. \tag{7.97b}$$

Hence, the least squares geoid height estimator becomes:

$$\hat{N} = \frac{1}{\gamma_0} \sum_{n=2}^{M} B_n \hat{k}_n T_{rr,n} + \frac{1}{\gamma_0} \sum_{n=0}^{\infty} B_n (1 - \kappa_n k_n) T_n \tag{7.98a}$$

with the expected MSE

$$m_{\hat{N}}^2 = \frac{1}{\gamma_0^2} \sum_{n=2}^{M} \frac{B_n^2 \sigma_{rr,n}^2 d\tau_n^2}{B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 d\tau_n^2} + \frac{1}{\gamma_0^2} \sum_{n=M+1}^{\infty} \frac{B_n^2 \sigma_{rr,n}^2 \tau_n^2}{B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 \tau_n^2}. \tag{7.98b}$$

Also this geoid height estimator can be written as an integral formula of the SGG data combined with the harmonic series of the EGM:

$$\hat{N} = \frac{R^2}{4\pi\gamma}\iint_{\sigma} F(\kappa,\psi)T_{rr}d\sigma + \frac{1}{\gamma_0}\sum_{n=2}^{M}\frac{B_n\sigma_{rr,n}^2}{B_n\sigma_{rr,n}^2 + \kappa_n^2\tau_n^2}T_n, \tag{7.99}$$

and the kernel function $F(\kappa,\psi)$ is formally given by Eq. (7.94b), but now with parameters $\hat{k}_n$ given by Eqs. (7.97a and 7.97b).

7.5.4 Spectral Combination of Data from Terrestrial Gravity, SGG and an EGM

We now assume that the terrestrial gravity anomaly and the SGG data (Δg and T_{rr}, respectively) have random observation errors with expectations zero and error degree variances σ_n^2 and $\sigma_{rr,n}^2$, and that they are available together with an EGM complete to degree M. A general geoid height estimator from these data three sets can be written:

$$\tilde{N} = \frac{1}{\gamma_0}\sum_{n=2}^{\infty}\left(C_n h_n \Delta g_n + B_n k_n T_{rr,n}\right) + \frac{1}{\gamma_0}\sum_{n=2}^{M} a_n T_n, \tag{7.100}$$

where B_n, $T_{rr,n}$ and T_n are the same symbols as in the previous sections and $C_n = R/(n-1)$. Moreover, h_n, k_n and a_n are arbitrary spectral weights. Requiring an unbiased estimator (with no dependence on the signal degree variances for any degree) yields the conditions

$$h_n + k_n\kappa_n + a_n - 1 = 0 \,\text{for}\, n \le M \tag{7.101a}$$

and

$$h_n + k_n\kappa_n - 1 = 0 \,\text{for}\, n > M, \tag{7.101b}$$

and by considering these conditions and eliminating k_n, the expected global mean-square error of $\tilde{N}$ becomes :

$$\begin{aligned} m_{\tilde{N}}^2 = {} & \frac{1}{\gamma^2}\sum_{n=2}^{M}\left[C_n^2 h_n^2\sigma_n^2 + B_n^2(h_n + a_n - 1)^2\kappa_n^{-2}\sigma_{rr,n}^2 + a_n^2 d\tau_n^2\right] \\ & + \frac{1}{\gamma_0^2}\sum_{n=M+1}^{\infty}\left[C_n^2 h_n^2\sigma_n^2 + B_n^2(h_n - 1)^2\kappa_n^{-2}\sigma_{rr,n}^2\right] \end{aligned} \tag{7.102}$$

Differentiating the last equation with respect to h_n and a_n and equating it to zero, one obtains the following two sets of equations:

(1) for $n \leq M$:

$$h_n\left(\kappa_n^2 C_n^2 \sigma_n^2 + B_n^2 \sigma_{rr,n}^2\right) + a_n B_n^2 \sigma_{rr,n}^2 = B_n^2 \sigma_{rr,n}^2 \tag{7.103a}$$

$$h_n B_n^2 \sigma_{rr,n}^2 + a_n\left(B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 d\tau_n^2\right) = B_n^2 \sigma_{rr,n}^2, \tag{7.103b}$$

and

(2) for $n > M$:

$$h_n\left(\kappa_n^2 C_n^2 \sigma_n^2 + B_n^2 \sigma_{rr,n}^2\right) = B_n^2 \sigma_{rr,n}^2. \tag{7.104}$$

The least squares solutions for the coefficients thus become:

$$\text{for } n \leq M: \ \hat{h}_n = \frac{B_n^2 \sigma_{rr,n}^2 dc_n}{G_n}, \ \hat{k}_n = \frac{C_n^2 \sigma_n^2 d\tau_n^2 \kappa_n}{G_n} \text{ and } \hat{a}_n = \frac{B_n^2 \sigma_{rr,n}^2 C_n^2 \sigma_n^2}{G_n}, \tag{7.105a}$$

where:

$$G_n = B_n^2 \sigma_{rr,n}^2 C_n^2 \sigma_n^2 + \kappa_n^2 C_n^2 \sigma_n^2 d\tau_n^2 + B_n^2 \sigma_{rr,n}^2 d\tau_n^2, \tag{7.105b}$$

and

$$\text{for } n > M: \quad \hat{h}_n = \frac{B_n^2 \sigma_{rr,n}^2}{G_n}, \quad \hat{k}_n = \frac{C_n^2 \sigma_n^2 \kappa_n}{G_n} \text{ and } \hat{a}_n = 0, \tag{7.106a}$$

where:

$$G_n = B_n^2 \sigma_{rr,n}^2 + \kappa_n^2 C_n^2 \sigma_n^2. \tag{7.106b}$$

Using these spectral weights, we arrive at the least squares solution for the geoid height either from Eq. (7.100) or, partly in the space domain, as:

$$\hat{N} = \frac{R}{4\pi\gamma_0} \iint_\sigma S(\kappa, \psi) \Delta g d\sigma + \frac{R}{4\pi\gamma_0} \iint_\sigma K(\kappa, \psi) T_{rr} d\sigma + \frac{1}{\gamma_0} \sum_{n=2}^{M} B_n \hat{a}_n T_n, \tag{7.107a}$$

where

$$S(\kappa,\psi)=\sum_{n=2}^{\infty}\frac{2n+1}{n-1}\hat{h}_n P_n(\cos\psi). \tag{7.107b}$$

Obviously, $S(\kappa,\psi)$ is a weighted form of Stokes' function, and

$$K(\kappa,\psi)=\sum_{n=2}^{\infty}\frac{2n+1}{(n+1)(n+2)}\hat{k}_n P_n(\cos\psi). \tag{7.107c}$$

Finally, by inserting Eqs. (7.105a), (7.105b), (7.106a) and (7.106b) into Eq. (7.102), one obtains the expected MSE of $\hat{N}$:

$$m_{\hat{N}}^2=\sum_{n=2}^{M}\frac{C_n^2\sigma_n^2B_n^2\sigma_{rr,n}^2d\tau_n^2}{G_n}+\sum_{n=M+1}^{\infty}\frac{C_n^2\sigma_n^2B_n^2\sigma_{rr,n}^2}{G_n}. \tag{7.108}$$

7.5.5 *Spectral Smoothing and Combination with Airborne Gravity Data*

Airborne gravity data can be applied similar to SGG data as discussed above. It has the advantage over satellite data that the altitude is lower, which results in a stronger signal and less smoothing needed for the downward continuation to sea level.

The downward continuation and smoothing of the gravity disturbance is given by Eqs. (7.78a and 7.78b) with the changes that $A_n=1$ and T_{rr} is replaced by the observation δg at a flight level of radius r. Then the predicted gravity disturbance at sea level becomes:

$$\delta\hat{g}^*=\sum_{n=0}^{\infty}\frac{c_{\delta g,n}^2}{F_n}\delta g_n, \tag{7.109a}$$

where

$$F_n=\alpha_n^2c_{\delta g,n}^2+\sigma_n^2;\quad \alpha_n=(R/r)^{n+2} \tag{7.109b}$$

and σ_n^2 is the error-degree variance of the gravity disturbance, and the expected MSE can be expressed as:

$$m_{\delta\hat{g}}^2=\sum_{n=0}^{\infty}\frac{c_{\delta g,n}^2\sigma_n^2\alpha_n^2}{F_n}. \tag{7.110}$$

Similarly the geoid height is predicted by:

$$\hat{N} = \frac{R}{\gamma_0}\sum_{n=0}^{\infty}\frac{\alpha_n c_{\delta g,n}^2}{(n+1)F_n}\delta g_n \tag{7.111a}$$

with the MSE

$$m_{\hat{N}}^2 = \left(\frac{R}{\gamma_0}\right)^2\sum_{n=0}^{\infty}\frac{\alpha_n^2 c_{\delta g,n}^2\sigma_n^2}{F_n}. \tag{7.111b}$$

In the spectral combination of airborne gravity disturbances with an EGM (represented by its disturbing potential Laplace harmonics T_n, complete to degree M), one may start from the general estimator

$$\tilde{N} = \frac{1}{\gamma_0}\sum_{n=0}^{\infty}K_n k_n \delta g_n + \frac{1}{\gamma_0}\sum_{n=0}^{M}a_n T_n \tag{7.112a}$$

where $K_n = R/(n+1)$, and

$$\varepsilon_{\tilde{N}} = \frac{1}{\gamma_0}\sum_{n=0}^{M}\left[K_n k_n \varepsilon_n + a_n \varepsilon_n^T + (k_n\kappa_n + a_n - 1)T_n\right] + \frac{1}{\gamma_0}\sum_{n=M+1}^{\infty}\left[K_n k_n \varepsilon_n + (k_n\kappa_n - 1)T_n\right]. \tag{7.112b}$$

Requiring that the estimator should be unbiased through degree M leads to the conditions

$$k_n\kappa_n + a_n - 1 = 0 \text{ for all } n \le M. \tag{7.113}$$

Comparing Eq. (7.112a) with Eq. (7.86), one notices that the only differences are that the satellite-gradiometry observation is now replaced by the airborne-gravity disturbance and the factor B_n is replaced by factor K_n, so that the unbiased least squares estimator of the geoid height now has its equivalence in Eq. (7.114a) with the mean square error given by Eq. (7.114b):

$$\hat{N} = \frac{1}{\gamma_0}\sum_{n=2}^{\infty}K_n \hat{k}_n T_{rr,n} + \frac{1}{\gamma_0}\sum_{n=2}^{M}K_n(1-\kappa_n k_n)T_n \tag{7.114a}$$

with the expected MSE

$$m_{\hat{N}}^2 = \frac{1}{\gamma_0^2}\sum_{n=2}^{M}\frac{K_n^2\sigma_n^2 d\tau_n^2}{K_n^2\sigma_n^2 + \kappa_n^2 d\tau_n^2} + \frac{1}{\gamma_0^2}\sum_{n=M+1}^{\infty}\frac{K_n^2\sigma_n^2\tau_n^2}{K_n^2\sigma_n^2 + \kappa_n^2\tau_n^2}. \tag{7.114b}$$

In a similar way, as in the previous subsection, this estimator can also be combined with terrestrial gravity data.

7.5.6 Concluding Remarks

Based on the numerical applications described in Sjöberg and Eshagh (2012), the following conclusions are drawn:

The least-squares solutions derived by spectral combination enable the optimum determination of the geoid height without prior time-consuming downward continuation of the satellite-gravity gradiometry (SGG) data. The ill-posed inverse problem for the downward continuation, usually formulated by an integral equation, whose solution needs regularization, is here replaced by an integral formula. The typical amplification of the propagated observation errors when performing the downward continuation is efficiently smoothed by the least-squares spectral filter provided in our technique. The theoretical mean-square error of the estimator is easily computed for *a priori* analyses or together with the actual estimates after computations. The kernel functions derived above cannot be written in closed forms, implying an essential computational burden for each of them. However, for their practical evaluations, they may be tabulated, and the actual values of each kernel, to be used in the numerical integrations, can be interpolated from the table. Numerical studies on the biased and unbiased types of spectral combination of SGG data and an Earth gravitational model show insignificant difference between the filtering kernels and their global root-mean-square errors. The contribution of far-zone data remains small in the integrals. In the combination of SGG data, EGM08 and terrestrial data, the kernel related to the gravity anomaly is similar to Stokes' original kernel, and it approaches zero at a geocentric angle of 3°, while the kernel involving the SGG data needs a larger coverage of the data; to a geocentric angle of about 5°.

Finally, a numerical example in Kern et al. (2003), including satellite data, provides too optimistic results, which disagree with our results, obviously due to a computational mistake of placing the satellite at sea level.

7.6 Applications of Atomic Clocks in Physical Geodesy

There are several effects from both special and general relativity applied in satellite geodesy. Bjerhammar (1975) and (1985) presented a technique to apply atomic-clock to determine potential differences. See also Vermeer (1983). It is well-known from general relativity that the time recorded by a clock depends on the gravitational potential. This means also that the frequencies f_P and f_Q measured by

two clocks at points P and Q are related to the geopotentials at the two points by the simple formula:

$$\frac{f_P^2}{f_Q^2} = \frac{1 - 2W_Q/c^2}{1 - 2W_P/c^2}, \tag{7.115}$$

which can be written:

$$W_P - W_Q = \left(\frac{f_P^2 - f_Q^2}{f_P^2}\right)\left(\frac{c^2}{2} - W_P\right) = \frac{\bar{f}\Delta f}{f_Q^2}c^2 + \varepsilon_W, \tag{7.116a}$$

where:

$$\varepsilon_W = -2W_P\frac{\bar{f}\Delta f}{f_Q^2} \approx -2\frac{\bar{f}\Delta f}{f_Q^2}\frac{GM}{r_P}, \tag{7.116b}$$

$\Delta f = f_P - f_Q$ and $\bar{f} = (f_P + f_Q)/2$. As $GM/r_P \leq R\gamma_0$, it follows that the residual term ε_W is only of the order of 4% of the main term on the right hand-side. Hence, to this order of approximation, one finally obtains the equation

$$W_P - W_Q \approx \frac{\bar{f}\Delta f}{f_Q^2}c^2 \approx \frac{\Delta f}{\bar{f}}c^2, \tag{7.117}$$

which can be used to directly determine the geopotential difference between two points on the Earth's surface or in space. To determine the potential difference corresponding to heights of the order of 1 cm, the atomic-clock frequency must be accurate to 10^{-18} s^{-1}, and this goal is now (2016) realistic in the very near future.

Consider that the above technique is used in a regional or global network to determine geopotential numbers of the network. If the geometric heights are also determined, e.g. by VLBI or GNSS, the normal heights can be fixed by Eq. (3.94a–3.94c), and the difference between geodetic heights and normal heights are the quasigeoid heights. Finally, geoid heights and orthometric heights can be estimated by the formulas given in Sect. 7.2. See also Chap. 9.

Appendix

Let the n-th degree radial component observation error be represented by the harmonic series

$$\varepsilon_{rr,n} = \sum_{m=-n}^{n} \varepsilon_{nm} Y_{nm}, \tag{7.118}$$

where $Y_{nm} = Y_{nm}(\theta, \lambda)$. Here (θ, λ) are the co-latitude and longitude of the spherical harmonic Y_{nm}. Taking the global average of two degree errors $\varepsilon_{rr,n}$ and $\varepsilon_{rr,p}$ with $n \neq p$, it follows from the orthogonality of the spherical harmonics when integrated over the sphere that:

$$q_{np} = \frac{1}{4\pi} \iint_{\sigma} \varepsilon_{rr,n} \varepsilon_{rr,p} d\sigma = 0. \tag{7.119}$$

Hence, the degree-correlations, given by $E\{q_{np}\}$, also vanish.

References

Ågren J (2004) The analytical continuation bias in geoid determination using potential coefficients and terrestrial gravity data. J Geod 78:314–332

Bjerhammar A (1962) Gravity reduction to a spherical surface. Royal Institute of Technology, Division of Geodesy, Stockholm

Bjerhammar A (1963) A new theory of gravimetric geodesy. Royal Institute of Technology, Division of Geodesy, Stockholm

Bjerhammar A (1973) Theory of errors and generalized matrix inverses. Elsevier Scientific Publ. Co, Amsterdam

Bjerhammar A (1975) Discrete approaches to the solution of the boundary value problems of physical geodesy. Boll Geod Szi Aff 34(2):185–240

Bjerhammar A (1985) On a relativistic geodesy. Bull Géod 59:207–220

Bruinsma SL, Marty JC, Balmino G, Biancale R, Förste C, Abrikosov O, Neumayer H (2010) GOCE gravity field recovery by means of the direct numerical method, presented at the ESA Living Planet Symposium, 27th June–2nd July 2010, Bergen, Norway

Dayoub N, Edwards SJ, More P (2012) The Gauss-Listing potential value W_0 and its rate from altimetric mean sea level and GRACE. J Geod 86:681–694

Flury J, Rummel R (2009) On the geoid-quasigeoid separation in mountain areas. J Geod 83:829–847

Fotopoulos G (2003) An analysis of the optimal combination geoid, orthometric and ellipsoidal height data, UCGE report 20185, Depth Geomat Eng, Univ Calgary

Fotopoulos G (2013) Combination of heights. In: Sanso F, Sideris MG (eds) Geoid determination: theory and methods, lecture notes in earth system science, SpringerLink, pp 517–544

Gauss FW (1828) Bestimmung des Breitenunterschiedes zwischen den Sternwarten von Göttingen und Altona durch Beobachtungen am Ramsdenschen Zenithsector, Vanderschoeck und Ruprecht, Göttingen, pp 48–50

Groten E (2004) Fundamental constants and their implications. Allg. Vermessungsnachr

Heiskanen W, Moritz H (1967) Physical geodesy. W. H. Freeman and company, San Francisco and London

Hotine M (1969) Mathematical geodesy, ESSA Monograph 2, U.S. Dept. of Commerce, Washington, D.C

Kern M, Schwarz KP, Sneeuw N (2003) A study on the combination of satellite, airborne and terrestrial gravity data. J Geod 77:217–225

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy, Meddelelse No. 44, Geodaetisk Institut, Copenhagen

Lauritzen SL (1973) The probabilistic background of some statistical methods in physical geodesy. Rep. No. 48, Geod. Inst., Copenhagen

Listing JB (1873) Über unsere jetzige Kenntnis der Gestalt und Grösse der Erde. Nachr. D Kgl Gesellschaft d Wiss und der Georg-August-Univ, Göttingen, pp 33–98
Martinec Z (1998) Boundary-value problems for gravimetric determination of a precise geoid. Lecture Notes in Earth Sciences 73. Springer, New York
Mather RS (1978) The earth's gravity field and ocean dynamics, NASA Tech. Memo. 79540, May 1978
Moritz H (1972) Advanced least-squares methods, Reports of the Department of Geodetic Science, Report No. 175, Ohio State University
Moritz H (1980) Advanced physical geodesy. Wichmann, Karlsruhe
Metzler B, Pail R (2005) GOCE data processing: the spherical cap regularization approach. Stud Geophys Geod 49:441–462
Migliaccio F, Reguzzoni M, Sansò F, Tscherning CC, Veicherts M (2010) GOCE data analysis: the space-wise approach and the first space-wise gravity field model. In: Proceedings of the ESA living planet symposium, 28 June–2 July 2010, Bergen, Norway
Molodensky MS, Eremeev VF, Yurkina MI (1962) Methods for study of the external gravitation field and figure of the earth. Transl. from Russian (1960), Jerusalem, Israel Program for Scientific Translations
Nicolls RJ, Cazenave A (2010) Sea-level rise and its impact on coastal zones. Science 328:1517–1520
Pail R, Goiginger H, Mayrhofer R, Schuh WD, Brockmann JM, Krasbutter I, Hoeck E, Fecher T (2010) GOCE gravity field model derived from orbit and gradiometry data applying the time-wise Method. In: Proceedings of the ESA living planet symposium, 28 June–2 July 2010, Bergen, Norway
Pavlis NA, Simon AH, Kenyon SC, Factor JK (2012) The development and evaluation of the earth gravitational model 2008 (EGM2008). JGR 117:B04406
Ries JC, Eans RJ, Shum CK, Watkins MM (1992) Progress in the determination of the gravitational coefficients of the Earth. GRL 19(6):529–531
Schuh WD, Brockmann JM, Kargoll B, Krasbutter I, Pail R (2010) Refinement of the stochastic model of GOCE scientific data and its effect on the in-situ gravity field solution. In: Proceedings of the ESA living planet symposium, 28 June–2 July 2010, Bergen, Norway
Sacerdote F, Sansó F (2004) Geodetic boundary value problems and the height datum problem. IAG Symposia, vol 127. Springer, New York, pp 174–178
Sanchez L (2012) Towards a vertical datum standardization under the umbrella of global geodetic observing system. J Geod Sci 2(4):325–342
Sjöberg LE (1975) On the discrete boundary value problem of physical geodesy with harmonic reductions to an internal sphere, The Royal Institute of Technology, Division of Geodesy, Stockholm
Sjöberg LE (1977) On the errors of spherical harmonic developments of gravity at the surface of the earth. Rep Dept Geod Sci No. 257, OSU, Columbus, Ohio
Sjöberg L (1978) A comparison of Bjerhammar's methods and collocation in physical geodesy Rep. 273, Dep. of Geod. Sci., Ohio State Univ., Columbus, 1978
Sjöberg LE (1983) Unbiased estimation of variance covariance components in condition adjustment with unknowns a MINQUE approach. Zeitschrift für Vermessungswesen, 108, 9, pp. 382 387, 1983
Sjöberg LE (1984) Non negative variance component estimation in linear adjustment models. Manuscripta Geod 9:247–280
Sjöberg LE (2001) Topographic and atmospheric corrections of the gravimetric geoid determination with special emphasis of the effects of degrees zero and one. J Geod 75:283–290
Sjöberg LE (2007) The topographic bias by analytical continuation in physical geodesy. J Geod 81:345–350
Sjöberg LE (2010) A strict formula for geoid-to-quasigeoid separation. J Geod 84:699–702
Sjöberg LE (2011) On the definition and realization of a global vertical datum. J Geod Sci 1(3):154–157

Sjöberg LE (2012) The geoid-to-quasigeoid difference using an arbitrary gravity reduction model. Stud Geophys Geod 56:929–933
Sjöberg LE (2013a) New solutions for the geoid potential W0 and the mean earth ellipsoid dimensions. J Geod Sci 3(4):258–365
Sjöberg LE (2013b) The geoid or quasigeoid- which reference surface should be preferred for a national height system? J Geod Sci 3:103–109
Sjöberg LE (2015) Rigorous geoid-from-quasigeoid corrections using gravity disturbances. J Geod Sci 5:115–118
Sjöberg LE, Eshagh M (2012) A theory on geoid modelling by spectral combination of data from satellite gradiometry, terrestrial gravity and an Earth gravitational model. Acta Geod Geoph Hung 47(1):1–16
Tscherning CC (2013) Geoid determination by 3D least-squares collocation. In: Sanso F, Sideris MG (eds) Geoid determination; theory and methods, 311–329, Lecture notes in earth system sciences 110, Springer-Verlag Berlin Heidelberg 2013, doi:10.1007/978-3-540-74700-0
Vermeer M (1983) Chronometric levelling. Rep Finnish Geod Inst 1983:1–7

Chapter 8
Gravity Inversion

Abstract This chapter describes gravity inversion for various geophysical, geodetic and geodynamic applications. After introducing some basic geophysical concepts and formulas for gravity inversion, various isostatic models and their use in determining crustal depth by gravity is described at length and compared and combined with seismic models of crustal depth. The preferred isostatic model is based on Vening Meinesz-Moritz hypothesis with a global isostatic compensation of Bouguer gravity disturbances rather than Bouguer gravity anomalies. There are also applications of gravity for estimating tectonic stress in the mantle and viscosity in the mantle (based on post-glacial land uplift rate data in Fennoscandia). Temporal changes of the gravity field by many years of monthly repeated data from the satellite mission GRACE are used to demonstrate their power in determining large-scale Earth mass and geoid changes, such as positive geoid height rates in Laurentia and Fennoscandia related to glacial isostatic adjustments, and negative rates in Greenland and West Antarctica, as a result of mass losses due to ice-sheet melting. Also, the upper mantle viscosity in Fennoscandia is determined by these data.

Keywords Crustal thickness · Gravity inversion · Moho · Tectonic stress · Temporal changes of gravity · Viscosity in the mantle

8.1 Introduction

Gravity inversion is an important tool for detecting, understanding and interpreting several kinds of geophysical structures and geodynamical processes. Many ongoing geodynamical spatial and temporal variations can be obtained from gravity data all over the Earth with several applications. The gravity field also gives information on the Earth's crustal evolution and isostatic compensation, as well as stress patterns in the lithosphere.

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_8

8.1.1 Basic Geophysical Concepts

The interior structure of the Earth is roughly layered in spherical shells. These layers can be defined by either their chemical or their rheological properties. Scientific understanding of the internal structure of the Earth is based on observations of topography and bathymetry, observations of rock in outcrop, samples brought to the surface from greater depths by volcanic activity, analysis of the seismic waves that pass through the Earth, measurements of the gravitational and magnetic fields of the Earth using terrestrial and satellite based sensors and experiments with crystalline solids at pressures and temperatures characteristics of the Earth's deep interior.

8.1.1.1 Compositional Layers

The structure of Earth can be defined in different ways, e.g. by well-known properties such as rheology or based on chemical properties. The Earth contains three main compositional layers (Skinner and Porter 1995):

- The densest of the three layers, which is located in the centre, is the inner core, largely composed of solid iron.
- Outside the inner core is the liquid outer core.
- The thick shell with rocky material that surrounds the core is called the mantle. It is less dense than the inner core but denser than the outermost layers.
- Outside the mantle lies the thinnest part of the Earth, the crust, which consists of solid rocky material that is less dense than the mantle.

The different layers have different thicknesses, volumes and densities, and the densities vary both laterally and with depth (see Fig. 8.1). The volumes are very disparate, and especially the volume of the mantle is very large in contrast to those of the others: crust 2%, mantle 80%, outer core 17%, and inner core 1% of the Earth's volume (Bott 1971, p. 10). The Earth's crust, composed of a variably solid material, is its hard outer shell which overlies the mantle. Compared to other layers, it is much thinner and floats upon the softer mantle. The crust can be divided into two main parts: oceanic and continental crust. The oceanic crust is about 6–12 km thick, consisting mainly of heavy rocks, like basalt with an average density of 3 g/cm^3, which is close to the upper mantle density. The continental crust can be divided into six big plates: Eurasia (Europe and Asia together), Africa, North America, South America, Antarctica and Australia. The crust is thicker below the continents, with an average of about 36 km and with a maximum of 80 km in Tibet.

One question that can be asked here is how we know about the properties and composition of the various layers. A likely answer can be found in Skinner and Porter (1995), and it is by the indirect measurements that are used today. One way to determine the composition is to measure how the density of rock changes with depth. It can be measured by the speed that earthquake waves pass through the

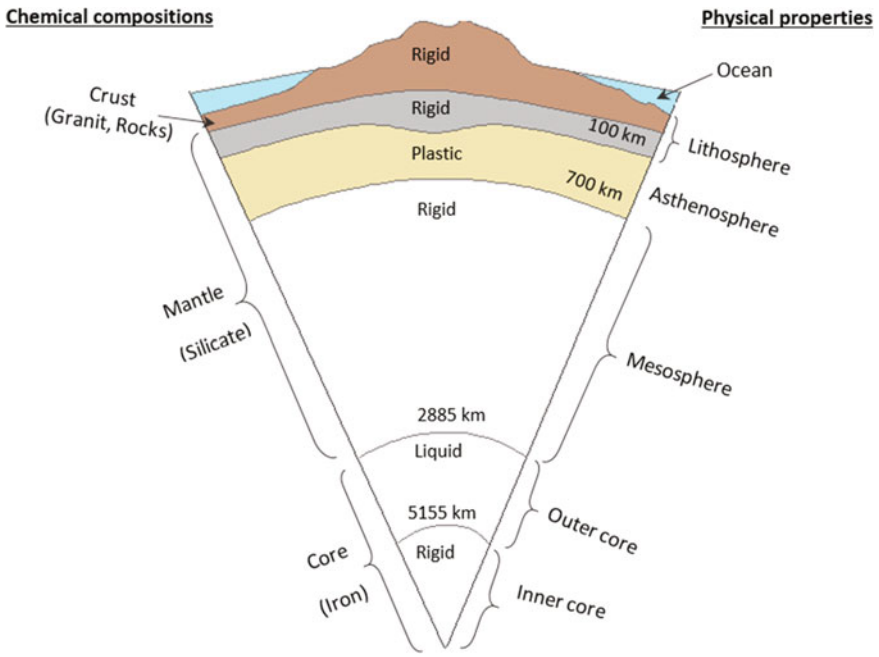

Fig. 8.1 A cross-sectional view of Earth showing Earth's layers classified by chemical composition along the *left side* of the diagram. For comparison, Earth's layers classified by physical properties are shown along the *right side* of the diagram

Earth's layers. The denser the rock, the faster propagates the waves. Due to sudden increases in velocity at various depths, one can infer that the solid Earth does not have a uniform composition, but must instead consists of layers with different densities.

8.1.1.2 Earth's Layers Based on Physical Properties

The Earth's layers can also be divided by their physical properties (such as rock strength and solid-versus-liquid condition), which roughly change with depth in the Earth's interior. It is evident that the changes in physical properties in the Earth's layers are controlled rather by temperature and pressure than by composition. The major layers where physical properties change are (Skinner and Porter 1995):

- between the inner and outer core.
- the mesosphere, which is a solid region of high temperature in the mantle that extends from the core-mantle boundary (at the depth of $\sim$2900 km) to a depth of about 350 km. The physical properties of this layer are characterized by its high temperature and pressure. It is important to mention that, when a solid is

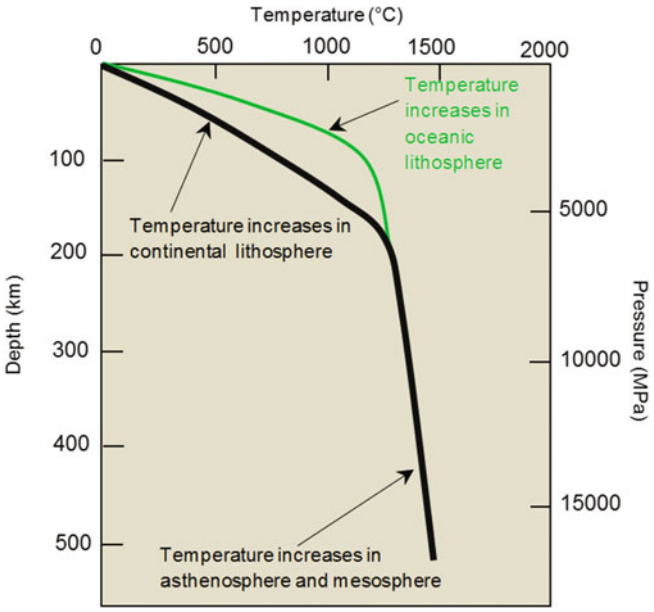

Fig. 8.2 Geothermal gradients (after Skinner and Porter 1995)

heated, it loses strength. Again, when it is compressed, it gains strength. The Mesosphere has relatively high strength.

- the asthenosphere, located 350 km to 100–200 km below the Earth's surface. It is a weak part of the mantle, because it is located where the balance between temperature and pressure is such that rocks have little strength and they easily deform.
- and finally the lithosphere: it is located above the asthenosphere (outermost strength zone) where rocks are cooler, stronger and more rigid. This is only because of differences in the temperature and pressure in rocky materials in the lithosphere and upper asthenosphere.

Figure 8.2 shows the relationship between depth, temperature and pressure in the Earth's interior. Geothermal gradients in the lithosphere are not constant. The largest and most important differences are results of the different thicknesses of oceanic and continental lithosphere. The figure shows that the temperature increases faster within the oceanic crust than the continental crust. The temperature changes gradually below the lithosphere, and it is determined by convection in the asthenosphere and mesosphere.

In Fig. 8.3, we present a global map representing the Earth's mass distribution as a surface-density distribution by Eq. (8.5b) and using the data from the GOCO-03S EGM (Mayer-Guerr et al. 2012) to degree and order 90. It shows very large mass deficiencies south of India, in NE Canada and in West Antarctica. The largest mass surpluses are found north of Australia, Chile and Tibet. Comparing Fig. 8.3 with a

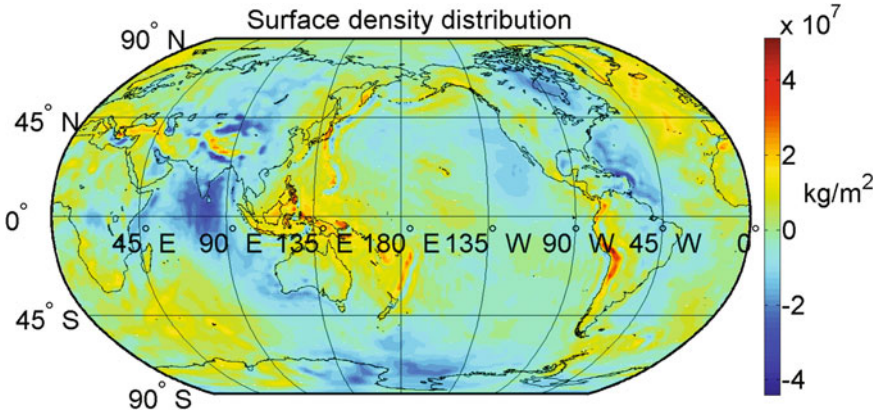

Fig. 8.3 Global map of the surface-density-distribution representation of Earth's mass distribution computed by Eq. (8.5b). Unit: kg/m^2

global geoid map (see Fig. 1.5) can explain the mass deficiencies and surpluses. Where a mass deficiency exists, the geoid surface will dip below the ellipsoid. Conversely, where there is a mass surplus, the geoid surface will rise above the ellipsoid. These influences cause the geoid surface to deviate from the normal ellipsoid within approximately ±100 m. The largest presently known undulations are the minimum in the Indian Ocean centred at 2° N and 80° E of −106.6 m and the maximum north of Australia of +87.6 m (according to the EGM2008 geoid model). These undulations are related to density distributions in the lower layers of the Earth. For example, the minimum geoid undulation (Fig. 1.5) is observed south of India in the Indian Ocean. Using spectral analysis, one finds large-wavelength sources at up to 1300 km of depth. This shows that there is a regional gravity anomaly with a negative density contrast at the level of 1300 km, where a sharp change in the gradient of the seismic velocities occurs. There is a low-viscosity zone below the transition zone in mantle discontinuity that may correspond to low-density rocks containing subducted rocks (Mishra and Ravi 2012). In contrast, the large negative-mass anomalies in West Antarctica and NE Canada are caused by mass deficiencies due to former ice sheets.

8.2 Basic Formulas in Inversion of Satellite Gravity-Field Models

Consider an EGM applied to the disturbing potential and the gravity anomaly, respectively (Sect. 3.2.4):

$$T_P = \frac{GM}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+1}\sum_{m=-n}^{n}C_{nm}Y_{nm}(P) = \sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+1}T_n(P) \tag{8.1}$$

and

$$\Delta g_P = \frac{GM}{R^2}\sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+2}(n-1)\sum_{m=-n}^{n}C_{nm}Y_{nm}(P) = \sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+2}\Delta g_n(P), \tag{8.2a}$$

where

$$M = \frac{4\pi}{3}\bar{\rho}R^3 \tag{8.2b}$$

is the mass of the Earth with $\bar{\rho}$ being the Earth's mean density.

To get a general picture of the lateral distribution of mass anomalies in the Earth, one may also represent the disturbing potential as a surface layer with density κ (in the unit of mass/surface element), yielding the surface integral on the sphere of radius R for $r_P = R$

$$T_P = GR^2\iint_{\sigma}\frac{\kappa}{l_P}d\sigma \tag{8.3}$$

In the spectral domain, Eq. (8.3) reads as:

$$T_P = GR\sum_{n=0}^{\infty}\frac{1}{2n+1}\sum_{m=-n}^{n}\kappa_{nm}Y_{nm}(P) \tag{8.4}$$

and, comparing the harmonic coefficients in Eqs. (8.1) for $r_P = R$ and (8.4), one obtains:

$$\kappa_{nm} = (2n+1)\frac{M}{R^2}C_{nm} = \frac{4\pi\bar{\rho}R}{3}(2n+1)C_{nm}, \tag{8.5a}$$

Hence, by summing up the harmonics, the surface-density distribution on the sphere can be estimated by the series

$$\kappa(P) = \frac{M}{R^2}\sum_{n=0}^{\infty}(2n+1)\sum_{m=-n}^{n}C_{nm}Y_{nm}(P) = \frac{4\pi\bar{\rho}R}{3}\sum_{n=0}^{\infty}(2n+1)\sum_{m=-n}^{n}C_{nm}Y_{nm}(P). \tag{8.5b}$$

In practice the series is always limited to a finite degree. This surface layer is a fictitious representation of the mass variations in the Earth. In reality, the relationship between the mass density and the EGM is more complicated as the

surface-mass load causes a deformation of the whole Earth body, which in turn changes the gravity field. The resulting relationship between the surface-mass density and the harmonic coefficients can therefore be expressed (e.g. Wahr et al. 1998) as:

$$\kappa(P) = \frac{M}{R^2}\sum_{n=0}^{\infty}\frac{(2n+1)}{1+k_n}\sum_{m=-n}^{n}C_{nm}Y_{nm}(P) = \frac{4\pi\bar{\rho}R}{3}\sum_{n=0}^{\infty}\frac{(2n+1)}{1+k_n}\sum_{m=-n}^{n}C_{nm}Y_{nm}(P), \tag{8.5c}$$

where k_n, the so-called load Love numbers, can be modelled based on some Earth model. (See, e.g. Farrell 1972; Sun and Sjöberg 1999). Equation (8.5c) is important in estimating a real surface load from an EGM or the temporal change of the former by a sequence in time of the latter. Also, assuming that the surface density is caused by a thin mass layer of density ρ_w and thickness h_w, its thickness can be estimated by $h_w = \kappa/\rho_w$, and a similar formula holds also for temporal changes of masses (Sect. 8.7.1).

Next we assume that a residual part δT of the disturbing potential is generated by a density contrast $\Delta\rho$ between the constant depth D_0 and $D_0 + \delta D$, where δD may vary. Assuming that the density contrast may change only laterally, δT will be given by the Newton integral

$$\delta T_P = G\iint_{\sigma}\Delta\rho\int_{R_{0-\delta D}}^{R_0}\frac{r^2 dr}{l_P}d\sigma, \tag{8.6}$$

where R is mean sea level radius and $R_0 = R - D_0$ or, in the spectral domain, after integration with respect to r:

$$\delta T_P = GR^2\sum_{n=0}^{\infty}\frac{1}{n+3}\left(\frac{R}{r_P}\right)^{n+1}\left(\frac{R_0}{R}\right)^{n+3}\iint_{\sigma}\frac{\Delta\rho}{n+3}\left[1-\left(1-\frac{\delta D}{R_0}\right)^{n+3}\right]P_n(\cos\psi)d\sigma, \tag{8.7a}$$

which can be rewritten after using the addition theorem for spherical harmonics of Eq. (2.49) to:

$$\delta T_P = GR^2\sum_{n=0}^{\infty}\frac{1}{(2n+1)(n+3)}\left(\frac{R}{r_P}\right)^{n+1}\left(\frac{R_0}{R}\right)^{n+3}\sum_{m=-n}^{n}Y_{nm}(P) \times\iint_{\sigma}\Delta\rho\left[1-\left(1-\frac{\delta D}{R_0}\right)^{n+3}\right]Y_{nm}d\sigma. \tag{8.7b}$$

Assuming that the deviation of δD is small versus D_0, one may expand the last bracket by a Taylor series

$$\frac{1}{n+3}\left[1-\left(1-\frac{\delta D}{R_0}\right)^{n+3}\right]=\frac{\delta D}{R_0}-\frac{n+2}{2}\left(\frac{\delta D}{R_0}\right)^2+\frac{(n+2)(n+1)}{6}\left(\frac{\delta D}{R_0}\right)^3+\cdots \tag{8.7c}$$

and, by inserting the series into Eq. (8.7b) and carrying out the integration one obtains:

$$\begin{aligned}\delta T_P=&\frac{4\pi GR^2}{R_0}\sum_{n=0}^{\infty}\frac{1}{2n+1}\left(\frac{R}{r_P}\right)^{n+1}\left(\frac{R_0}{R}\right)^{n+3}\\&\times\sum_{m=-n}^{n}\left[(\delta D\Delta\rho)_{nm}-\frac{n+2}{2R_0}\left(\delta D^2\Delta\rho\right)_{nm}+\frac{(n+2)(n+1)}{6R_0^2}\left(\delta D^3\Delta\rho\right)_{nm}-\cdots\right]Y_{nm}(P),\end{aligned} \tag{8.8}$$

where $(\delta D\Delta\rho)_{nm}$ are the spherical harmonic coefficient of δD times $\Delta\rho$. Comparing the spectrum of this series with that of δT represented by Eq. (8.1),

$$\delta T_P=\frac{GM}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+1}\sum_{m=-n}^{n}\delta C_{nm}Y_{nm}(P)=\sum_{n=0}^{\infty}\left(\frac{R}{r_P}\right)^{n+1}\delta T_n(P), \tag{8.9}$$

where δT_n is the n-th Laplace harmonic of δT, and neglecting higher-order terms in the right member of Eq. (8.8), one obtains a first-order solution by:

$$\delta D\Delta\rho=\frac{1}{4\pi GR}\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{R_0}\right)^{n+2}\delta T_n(P). \tag{8.10a}$$

or, upon considering Eq. (8.2b) and the last part of Eq. (8.9):

$$\begin{aligned}\delta D\Delta\rho&=\sum_{n=0}^{\infty}a_n(P)=\sum_{n=0}^{\infty}\sum_{m=-n}^{n}a_{nm}Y_{nm}(P)\\&=\frac{\bar{\rho}R}{3}\sum_{n=0}^{\infty}(2n+1)\left(\frac{R}{R_0}\right)^{n+2}\sum_{m=-n}^{n}\delta C_{nm}Y_{nm}(P).\end{aligned} \tag{8.10b}$$

It is obvious that (8.10a, 8.10b) have to be truncated at some maximum degree, say $n_{\max}$, as the ratio $R/R_0>1$, which implies also that the series will be more and more sensitive to errors in δT_n for higher degrees.

There is no way to determine both the geometry (δD) and density ($\Delta\rho$) of a geological feature from gravity data alone. However, if $\Delta\rho$ is known (not necessarily laterally constant), the depth variation δD can thus be estimated to first order by:

$$\delta D = \frac{1}{\Delta\rho}\sum_{n=0}^{n_{\max}} a_n, \tag{8.11a}$$

and this formula can be improved iteratively to higher orders as follows. By summing up the terms and rearranging Eq. (8.8), one obtains:

$$\delta D_P = \frac{1}{\Delta\rho_P}\sum_{n=0}^{n_{\max}} a_n + \frac{1}{\Delta\rho_P}\sum_{n=0}^{n_{\max}}\left[\frac{n+2}{2R_0}\left(\delta D^2\Delta\rho\right)_n - \frac{(n+2)(n+1)}{6R_0^2}\left(\delta D^3\Delta\rho\right)_n + \cdots\right], \tag{8.11b}$$

where the higher-order harmonics can be determined iteratively by harmonic analysis of:

$$\delta D^2\Delta\rho \approx \frac{\left(\sum_{k=0}^{n_{\max}} a_k\right)^2}{\Delta\rho} \quad \text{and} \quad \delta D^3\Delta\rho^2 \approx \frac{\left(\sum_{k=0}^{n_{\max}} a_k\right)^3}{\Delta\rho}. \tag{8.12}$$

Alternatively, if δD is known, one obtains a first-order solution of the lateral density contrast from Eq. (8.10b) by:

$$\Delta\rho_P \approx \frac{R\bar{\rho}}{3(\delta D)_P}\sum_{n=0}^{n_{\max}}(2n+1)\left(1-\frac{D_0}{R}\right)^{-(n+2)}\sum_{m=-n}^{n}\delta C_{nm}Y_{nm}(P). \tag{8.13}$$

Note that in (8.13) the factor $1 - D_0/R$ increases with degree n, which makes the solution sensitive to errors in the data (δT) for higher degrees. This property is typical in solving inverse problems, and similar problems occur in all solutions with downward continuation as above. This negative effect can be diminished by some kind of *smoothing* or *regularization* (see Sect. 2.8) of the solution. Another smoothing operation is to remove short-wavelength signals and errors in the data. For example, smoothing is obtained by removing the effects of the atmosphere and topography prior to dwc (i.e. applying the DITE) and restoring the effects after downward continuation (PITE).

If the density contrast $\Delta\rho$ is known, one may improve the solution of δD by rewriting the integral in Eq. (8.6) as two integrals:

$$\delta T_P = G\iint_{\sigma}\Delta\rho\int_{R-D}^{R}\frac{r^2dr}{l_P}d\sigma - G\iint_{\sigma}\Delta\rho\int_{R_0}^{R}\frac{r^2dr}{l_P}d\sigma = DT_P - DT_{0P}, \tag{8.14}$$

where the second integral is a known potential that can be explicitly expressed as:

$$DT_{0P} = 4\pi G \sum_{n=0}^{\infty} \Delta\rho_n(P) \frac{R^{n+3} - R_0^{n+3}}{(2n+1)(n+3)r_P^{n+1}}, \tag{8.15}$$

In particular, if $\Delta\rho$ is constant, DT_{0P} is also a constant and given by:

$$DT_{0P} = 4\pi G\Delta\rho\left(R^3 - R_0^3\right)/(3r_P) = \frac{4\pi G\Delta\rho R^2}{r_P}\left(D_0 - \frac{D_0^2}{R} + \frac{D_0^3}{3R^2}\right). \tag{8.16}$$

As a result, Eq. (8.8) becomes:

$$\delta T_P = 4\pi GR \sum_{n=0}^{\infty} \frac{1}{2n+1}\left(\frac{R}{r_P}\right)^{n+1} \times \sum_{m=-n}^{n}\left[(D\Delta\rho)_{nm} - \frac{n+2}{2R}\left(D^2\Delta\rho\right)_{nm} + \frac{(n+2)(n+1)}{6R^2}\left(D^3\Delta\rho\right)_{nm} - \cdots\right] Y_{nm}(P) - DT_{0P}. \tag{8.17}$$

Equation (8.17) is valid for any point on the sphere. Hence, by comparing the spectral terms of Eqs. (8.9) and (8.17), one may identify each harmonic as an equation, and, by rearranging terms and summing up, one obtains for $r_P = R$:

$$D_P = b_P + \frac{1}{\Delta\rho_P}\sum_{n=0}^{\infty}\left[\frac{n+2}{2R}\left(D^2\Delta\rho\right)_n - \frac{(n+2)(n+1)}{6R^2}\left(D^3\Delta\rho\right)_n + \cdots\right], \tag{8.18a}$$

where the first-order solution is given by:

$$b_P = \frac{R\bar{\rho}}{3\Delta\rho_P}\sum_{n=0}^{\infty}(2n+1)\sum_{m=-n}^{n}\delta C_{nm}\mathrm{Y}_{nm}(P) + \frac{\left(DT_{0P}\right)_{r_P=R}}{4\pi G\Delta\rho_P}, \tag{8.18b}$$

and the higher-order harmonics can be determined approximately as above. If $\Delta\rho$ is constant, the solution simplifies to:

$$D_P = b_P + \frac{1}{\Delta\rho}\sum_{n=0}^{\infty}\left[\frac{n+2}{2R}\left(D^2\right)_n - \frac{(n+2)(n+1)}{6R^2}\left(D^3\right)_n + \cdots\right], \tag{8.19a}$$

with:

$$b_P = \frac{R\bar{\rho}}{3\Delta\rho}\sum_{n=0}^{\infty}(2n+1)\sum_{m=-n}^{n}\delta C_{nm}Y_{nm}(P) + D_0 - \frac{D_0^2}{R} + \frac{D_0^3}{3R^2}. \tag{8.19b}$$

It is interesting that this solution avoids the regularization problem of Eqs. (8.10a)–(8.13).

Once the anomalies of mass per unit area on the unit sphere $(\delta D \Delta\rho)$ have been determined, the total mass anomaly (δM) can be estimated by the integral

$$\delta M = \iint_{\sigma} \Delta\rho \int_{R_0-\delta D}^{R_0} r^2 dr d\sigma = R_0^2 \iint_{\sigma} \Delta\rho \left(\frac{1-(1-\delta D/R_0)^3}{3} \right) d\sigma, \tag{8.20}$$

where the surface integral can be limited to a region bounding the mass anomaly.

8.2.1 Regional Studies

For regional studies, spatial solutions can be more attractive than spectral solutions, and more so the more local the phenomenon is and/or the higher the requested resolution is. In this case, by assuming that the density contrast is constant, Eq. (8.14) and its radial derivatives can be written on the compact forms

$$RG \iint_{\sigma} \Delta\rho \begin{pmatrix} F(r_P,\psi,D) \\ F'(r_P,\psi,D) \\ F''(r_P,\psi,D) \end{pmatrix} d\sigma = \begin{pmatrix} \delta T(P) + DT_{0P} \\ \delta T'(P) + DT'_{0P} \\ \delta T''(P) + DT''_{0P} \end{pmatrix} \tag{8.21}$$

where the kernel functions are:

$$F(r_P,\psi,D) = \int_{R-D}^{R} \frac{r^2}{l_P} dr = \left[\frac{r+3r_P t}{2} l_P + P_2(t) \ln \Psi \right]_{r=R-D}^{R} \tag{8.22a}$$

$$F'(r_P,\psi,D) = \frac{1}{2} \left[3tl_P + (r+3r_P t) \frac{(r_P - rt)}{l_P} + 2P_2(t) r_P^2 \frac{r - rt - tl_P}{\Psi l_P} + 2r_P P_2(t) \ln \Psi \right]_{r=R-D}^{R}, \tag{8.22b}$$

where $t = \cos\psi$ and $\Psi = (r - r_P t + l_P)$ and

$$F''(r_P,\psi,D) = \frac{dF'()}{dr_P}. \tag{8.22c}$$

Equations (8.21) are Fredholm integral equations of the first kind, typical in improperly-posed problems of gravity inversion. If the unknown is the density contrast, the equations are linear, and, if the depth variation (δD) is sought, the equations are non-linear integral equations. The equations are somewhat stabilized by assuming that all atmospheric and topographic effects are removed (which correspond to Bouguer types of data), and the data are preferably downward continued to mean sea level (implying $r_P = R$). Although the integrals are strictly global, the

kernel functions taper off with geocentric radius (ψ), which justifies truncation of the integrals at some cap size. This is particularly the case for Eq. (8.22c).

If the density contrast distribution $\Delta\rho$ is known, one obtains:

$$DT_{0P} = 4\pi GR^2 \sum_{n=0}^{\infty} \frac{\Delta\rho_n(P)}{(2n+1)(n+3)} \left(\frac{R}{r_P}\right)^{n+1} \left[1 - \left(\frac{D_0}{R}\right)^{n+3}\right] \tag{8.23a}$$

$$DT'_{0P} = -4\pi GR \sum_{n=0}^{\infty} \frac{(n+1)\Delta\rho_n(P)}{(2n+1)(n+3)} \left(\frac{R}{r_P}\right)^{n+2} \left[1 - \left(\frac{D_0}{R}\right)^{n+3}\right] \tag{8.23b}$$

and

$$DT''_{0P} = 4\pi G \sum_{n=0}^{\infty} \frac{(n+1)(n+2)\Delta\rho_n(P)}{(2n+1)(n+3)} \left(\frac{R}{r_P}\right)^{n+3} \left[1 - \left(\frac{D_0}{R}\right)^{n+3}\right]. \tag{8.23c}$$

If the density contrast is constant, the equations simplify to:

$$R \iint_{\sigma} \begin{pmatrix} F(r_P,\psi,D) \\ F'(r_P,\psi,D) \\ F''(r_P,\psi,D) \end{pmatrix} d\sigma = \frac{1}{k} \begin{pmatrix} \delta T(P) + DT_{0P} \\ \delta T'(P) + DT'_{0P} \\ \delta T''(P) + DT''_{0P} \end{pmatrix}; \quad k = G\Delta\rho, \tag{8.24}$$

where:

$$\frac{1}{k} \begin{pmatrix} DT_{0P} \\ DT'_{0P} \\ DT''_{0P} \end{pmatrix} = \frac{4\pi}{3} R^3 \left[1 - \left(1 - \frac{D_0}{R}\right)^3\right] \begin{pmatrix} r_P^{-1} \\ -r_P^{-2} \\ 2r_P^{-3} \end{pmatrix} \tag{8.25}$$

8.2.2 Determination of Simple Mass Structures

The structure of the Earth's interior is rather well-known from seismology. In this section, some formulas for estimating some simple (spherical) structures in the Earth are presented. These can be expanded to other forms of structures, e.g. by formulas presented by Dobrin (1976) for geophysical prospecting.

Assume that a spherical body of mass anomaly dM is located at depth D below sea level. Also consider that gravity (disturbance, disturbing potential and/or radial derivative of gravity disturbance) are observed at height $H \geq 0$ above sea level. Then it holds that:

$$\text{Disturbing potential (anomaly)} \quad \delta T = GdM/(\mathrm{H} + \mathrm{D}) \tag{8.26a}$$

$$\text{Gravity disturbance} \quad \delta g = GdM/(H+D)^2 \tag{8.26b}$$

$$\text{Gravity gradient (vertical)} \quad \delta g_r = -2GdM/(H+D)^3, \tag{8.26c}$$

and these equations one can estimate the mass anomaly of the spherical structure as:

$$dM = \frac{(\delta T)^2}{G|\delta g|} \tag{8.27a}$$

or

$$dM = \frac{4(\delta g)^3}{G(\delta g_r)^2} \tag{8.27b}$$

The depth to the structure can be obtained from:

$$D = \frac{\delta T}{\delta g} - H \tag{8.28a}$$

or

$$D = \sqrt{-2\delta T/\delta g_r} - H \tag{8.28b}$$

The above simple formulas for estimating the depths of large mass anomalies frequently do not provide reasonable results mainly because the highs and lows of the geoid and gravity do not agree in their positions, and gravity gradients are much too local to be of interest in this context. For example, extreme gravity anomalies of the order of ± 400 mGal, occur along island arc and trench combinations in narrow pairs of bands of local mass surplus and deficiencies caused by volcanism and plate subduction, while integration of such a rough gravity anomaly field by Stokes' formula contributes/averages to nearly zero in the geoid undulation. These facts show that gravity alone cannot solve both the shape, magnitude and location of a density contrast in the Earth's interior, but other data, e.g. from seismic tomography, is needed to solve the general problem. Nevertheless, the longest wavelengths of the gravity potential field of spherical harmonic degrees n = 2 and 3 can be expected to be related with the mass anomalies of the core/mantle boundary at the depth of 2900 km, while wavelengths of degrees $4 \leq n \leq 10$, having much less power, are likely caused by mantle convection and plate tectonics in the lower and medium mantle, and for $n > 10$ the origin is mainly in the upper mantle and the crust. For more information, see Bowin (1994) and its references.

Table 8.1 shows an overview of the Earth's mass structures with each layer's contribution to the total mass of the Earth (5.97×10^{24} kg). For example the mantle extends from the base of the crust to a depth of about 2900 km. Except for the asthenosphere, the mantle is solid (Fig. 8.1), and its density, increasing with

Table 8.1 Earth's layers parameters (depth and mass, Anderson 1989)

Earth's layer	Depth [km]	Fraction of total Earth mass
Continental crust	0–80	0.004
Oceanic crust	0–13	0.001
Upper mantle	23–400	0.103
Transition zone	400–650	0.075
Lower Mantle	650–2900	0.492
Outer core	2900–5200	0.308

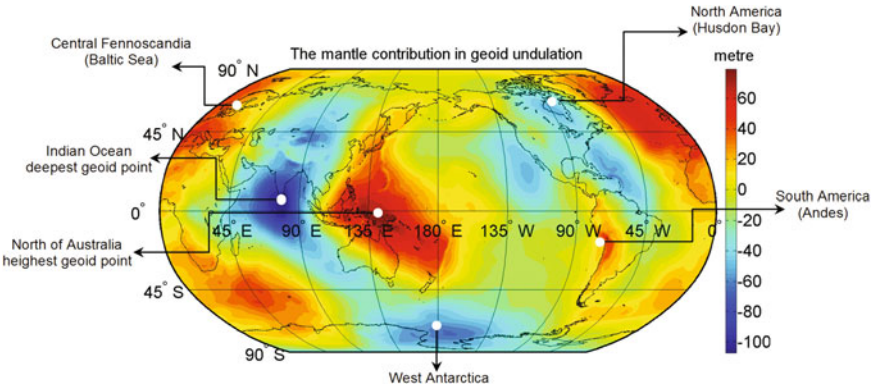

Fig. 8.4 Major global geoid undulations due to mantle-mass heterogeneities (using harmonic window $5 \leq n \leq 70$). Unit: metre

depth, ranges from 3.3 g/cm^3 (in the upper mantle) to about 6 g/cm^3 (in the lower mantle). The total mantle mass is 4×10^{24} kg, and its volume is 9.199×10^{20} m^3.

Applying Bruns' formula in Eq. (8.26a), the major global geoid disturbances due (only) to mantle mass are approximated in Fig. 8.4.

The figure shows that the contribution of the mantle mass is very significant to the geoid height. Large-scale disturbances are generated in the mantle (and also very deep and below the mantle for the longest wavelengths not shown in the figure), and small disturbances are due to mass variations in the Earth's crust. Swieczak et al. (2009) studied the geoid undulations and suggested that the causative sources for geoid anomalies are primarily located in the mantle. In fact, compensation of lithospheric loads related to large-wavelength anomalies occur in the mantle. The mantle mass is about 68.4% of the Earth's mass and 49.5% of the Earth's volume. The positions of the global extremes of geoid undulations in the Indian Ocean and NE of Australia are shown in Fig. 8.4, but here their magnitudes are less than those shown in the absolute geoid map of Fig. 1.5. This type of map can be used in studying mass anomalies in the solid Earth, e.g. for prospecting mineral deposits and determining the structure of the Earth.

8.3 Bouguer, No-Topography and Isostatic Gravity Anomalies and Disturbances

Geodesists' and geophysicists' definitions of the Bouguer gravity anomaly differ (e.g. Li and Götze 2001; Hackney and Featherstone 2003; Vajda et al. 2007). In geodesy, a gravity disturbance is the actual gravity minus normal gravity at the same point (P), while a modern gravity anomaly is gravity at P minus normal gravity at point Q at normal height along the normal to the reference ellipsoid through P. (At normal height means that the normal potential at Q equals the Earth's gravity potential at the observation point P.) In modern geodetic terminology, point P may be located anywhere in space, and particularly, when P is at the Earth's surface, one names the quantities surface-gravity disturbance and surface-gravity anomaly. [On the contrary, the classical gravity anomaly is only defined (with P) on the geoid (and Q on the reference ellipsoid).] The Bouguer gravity disturbance and anomaly means that the attraction of the topography has been removed.

In modern geophysics, the Bouguer gravity anomaly is the geodetic gravity disturbance (e.g. Hinze et al. 2005; Vajda et al. 2007). To avoid misunderstandings among and between geodesists and geophysicists, these two concepts (gravity anomaly and gravity disturbance) should not be mixed. In the presentation to follow, we use the modern geodetic terminology. So, at this point, we repeat the basic differences between a gravity anomaly (Δg) and a gravity disturbance (δg) at an arbitrary point P:

$$\Delta g_P = g_P - \gamma_Q \quad \text{and} \quad \delta g_P = g_P - \gamma_P, \tag{8.29}$$

where point Q is located at normal height along the normal to the reference ellipsoid through P. (Note that this is the common definition among geodesists, while the geophysical definition may vary.) For more information, see, e.g. Sjöberg et al. (2015).

Traditionally, gravity observations are reduced to free-air and Bouguer gravity anomalies using standard methods as discussed in Chap. 3. Traditionally, the Bouguer and isostatic gravity anomalies play the role of being the primary observables for gravity inversion of the Earth's interior and below-crust density structures. This is because the refined Bouguer anomaly is assumed to remove the gravity signal of the topography and crust, making the remaining gravity signal transparent to density features below the crust. However, as already emphasized in Sect. 4.2, the topographic reduction of the Bouguer anomaly is not complete, but it needs also the secondary indirect topographic effect (SITE) to be totally reduced for the topographic signal, leading to the so-called no-topography gravity anomaly. Usually it is assumed that the secondary indirect effect is small compared to the direct topographic effect. At least, that is the common experience in geoid determination. However, this effect may contribute most significantly, in particular, to long-wavelength topographic features in gravity inversion, which is the emphasis here. In the same way, the isostatic gravity anomaly yields an incomplete reduction

for the topography as well as for the isostatic compensation, properties that are achieved only for the refined isostatic gravity anomaly. This implies that the Vening Meinesz' isostatic equilibrium condition of a vanishing isostatic gravity anomaly holds only for the refined isostatic anomaly.

As in Chap. 3, we let V^T and its (negative) radial derivative (A^T) define the topographic potential and attraction, respectively, furthermore let V_c and A_c define the isostatic compensation potential and attraction, respectively. Then the traditional Bouguer gravity anomaly and gravity disturbance are given by Heiskanen and Moritz (1967, Sect. 3.3)

$$\Delta g^B = \Delta g - A^T \tag{8.30a}$$

and

$$\delta g^B = \delta g - A^T, \tag{8.30b}$$

where Δg and δg are the gravity anomaly and gravity disturbance, respectively, and $-A^T$ is *the direct effect on gravity*. Accordingly, the traditional isostatic gravity anomaly and gravity disturbance follow from:

$$\Delta g^I = \Delta g^B + A_C \tag{8.31a}$$

and

$$\delta g^I = \delta g^B + A_C. \tag{8.31b}$$

Here the total topographic signal is removed only for the Bouguer gravity disturbance, and part of it remains in the Bouguer anomaly. Therefore, for different applications of gravity inversion, there is a need for a new gravity anomaly. As seen in Eqs. (8.30a, 8.31b), the Bouguer gravity anomaly and disturbance experience the same topographic effects. Similarly, the traditional isostatic gravity anomaly and disturbance have the same topographic and compensation effects (see Eqs. 8.31a, 8.31b). However, from the fundamental equation of physical geodesy, Eq. (3.23a), we learn that:

$$\Delta g = \delta g + \frac{\partial \gamma}{\partial h}\frac{T}{\gamma} \approx \delta g - 2\frac{T}{r} \tag{8.32}$$

Accordingly, the disturbing potential T can be replaced with topographical potential $T = T^{nt} + V^T$, and it follows then (Sjöberg 1994, 1997, 2000; Vajda et al. 2007) that:

$$\Delta g = \Delta g^{nt} + \Delta g^T, \tag{8.33}$$

where Δg^{nt} is the no-topography gravity anomaly, and

$$\Delta g^T = -A^T + \frac{\partial \gamma}{\partial h}\frac{V^T}{\gamma} \approx -A^T + 2\frac{V^T}{r}, \tag{8.34}$$

i.e. the DITE of the gravity anomaly, which differs from that of the gravity disturbance ($-A^T$) by the term $2V^T/r$ and which is the secondarily indirect topographic effect (SITE). In other words, the Bouguer gravity anomaly, as defined by Eq. (8.30a), contains a remaining topographic contribution that is only removed by applying the SITE, while *the no-topography gravity anomaly*

$$\Delta g^{nt} = \Delta g - \Delta g^T \tag{8.35}$$

is consistent with the Bouguer gravity disturbance, δg^B, with no remnant attraction from the topography.

Similarly, the traditional isostatic gravity anomaly, as just cited, does not fully compensate for a topographic attraction, but it also needs to be corrected in accord with Eq. (8.34) to become *the refined isostatic gravity anomaly* (Sjöberg 2013)

$$\Delta g^{I,refined} = \Delta g^{nt} + A_C + \frac{\partial \gamma}{\partial h}\frac{V_C}{\gamma} \approx \Delta g^{nt} + A_C - 2\frac{V_C}{r}, \tag{8.36}$$

where the term $-2V_C/r$ is a remnant isostatic compensation. Only by adding this term does the isostatic gravity anomaly become consistent with the isostatic gravity disturbance:

$$\delta g^I = \delta g^B + A_C. \tag{8.37}$$

The detailed formulas for computing the topographic potentials and attractions of the spherical cylinder (Sjöberg 2007) and its isostatic compensation by Airy-Heiskanen's compensation model (see Sect. 8.4) are presented below.

Figure 8.5 shows a topographic model in the form of a spherical cylinder. Considering this figure, the topographic potential V^T at any point P of radius r_P along its axis is given by:

$$V^T(P) = 2\pi\mu \int_0^{\psi_0}\int_R^{r_s} \frac{r^2 dr\, \sin\psi d\psi}{l_{P0}} = \frac{2\pi\mu}{r_P}\int_R^{r_s} \left(l_{P0} - |r_P - r|\right) r dr \tag{8.38a}$$

where $\mu = G\rho$, $l_{P0} = \sqrt{r_P^2 + r^2 - 2r_P r\cos\psi_0}$. Equation (8.38a) can be integrated to:

$$V^T(P) = \frac{2\pi\mu}{r_P}\left(I(r_P, R, r_S) - r_P\frac{r_s^2 - R^2}{2} + \frac{r_s^3 - R^3}{3}\right) \quad \text{if} \quad r_P \geq r_s, \tag{8.38b}$$

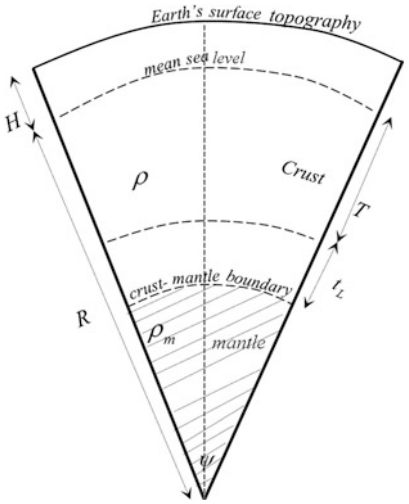

Fig. 8.5 A spherical cylinder and its isostatic compensation by Airy-Heiskanen's model

where:

$$\begin{aligned} I(r_P, R, r_S) &= \int_R^{r_s} l_{P0} r dr \\ &= \left[\frac{l_{P0}^3}{3} + r_P t_0 \left\{ \frac{r - r_P t_0}{2} l_{P0} + \frac{r_P^2}{2} \right.\right. \\ &\quad \left.\left. \times \left(1 - t_0^2\right) \ln 2(r - r_P t_0 + l_{P0}) \right\} \right]_{r=R}^{r=r_s} \end{aligned} \tag{8.38c}$$

where $t_0 = \cos \psi_0$.

Generally, the topographic gravity anomaly and gravity disturbance effects (corrections for removal of topographic signal) become (Eq. 5.4):

$$\Delta g^T = \delta g^T + 2\frac{V^T}{r_P} \quad \text{and} \quad \delta g^T = -A^T = \frac{\partial V^T}{\partial r_P}. \tag{8.39}$$

For the spherical cylinder, one obtains:

$$A^T = -\delta g^T = \frac{2\pi\mu}{r_P^2} \left(I(r_P, R, r_S) - r_P I'(r_P, R, r_S) + \frac{r_s^3 - R^3}{3} \right), \tag{8.40}$$

where

$$I'(r_P, R, r_S) = J_1(r_P, R, r_S) - J_2(r_P, R, r_S) \tag{8.41a}$$

with

$$J_1(r_P, R, r_S) = r_P \int_R^{r_s} \frac{r}{l_{P0}} dr = \left[l_{P0} + r_P t_0 \ln\left| 2r - 2r_P t_0 + 2l_{p0} \right| \right]_{r=R}^{r=r_S} \tag{8.41b}$$

and

$$\begin{aligned} J_2(r_P, R, r_S) &= -t_0 \int_R^{r_s} \frac{r^2}{l_{P0}} dr \\ &= \left[\frac{r + 3r_P t_0}{2} l_{P0} + \frac{3r_P^2 t_0^2 - r_P^2}{2} \ln\left| -2r_P t_0 + 2r - 2l_{P0} \right| \right]_{r=R}^{r=r_S} \end{aligned} \tag{8.41c}$$

The compensation potential for Airy-Heiskanen's isostatic model applied to the spherical cylinder is given by the integral

$$V_C(P) = 2\pi G \Delta\rho \int_0^{\psi_0} \int_{R-(D+t_L)}^{R-D} \frac{r^2 dr \, \sin\psi d\psi}{l_{P0}} \tag{8.42a}$$

where $t_L = 4.45H$, H is the topographic height, D denotes on crustal thickness and $\Delta\rho$ is the density contrast between the crust and mantle (Heiskanen and Moritz 1967, pp. 135–136). The integral can be evaluated to:

$$\begin{aligned} V_C(P) = \frac{2\pi G \Delta\rho}{r_P} \Bigg(& I(r_P, R - D - t_L, R - D) - r_P \frac{(R-D)^2 - (R - (D + t_L))^2}{2} \\ & + \frac{(R-D)^3 - (R - D - t_L)^3}{3} \Bigg), \quad \text{if} \quad r_P \geq r_s, \end{aligned} \tag{8.42b}$$

and the compensation attraction becomes:

$$\begin{aligned} A_C &= -\frac{\partial V_C}{\partial r_P} \\ &= \frac{2\pi G \Delta\rho}{r_P^2} \left(I(r_P, R - D - r_S, R - D) - r_P I'(r_P, R - D - r_S, R - D) \right. \\ &\quad \left. + \frac{(R-D)^3 - (R - D - t_L)^3}{3} \right). \end{aligned} \tag{8.42c}$$

Figures 8.6 and 8.7 show how the terms SITE = $\frac{2V^T}{r}$ and DITE = A^T of Eq. (8.42a) increase with H and ψ.

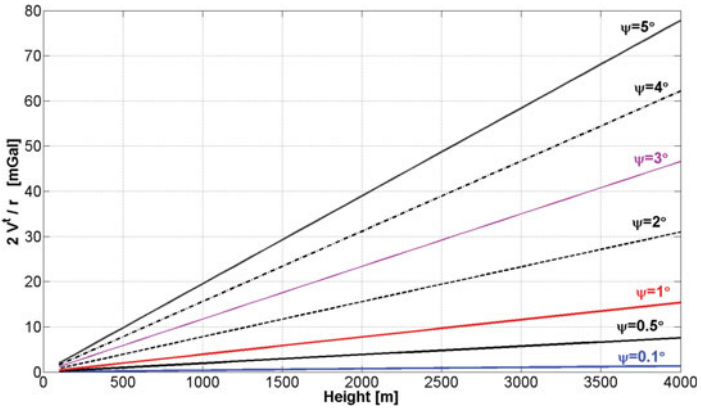

Fig. 8.6 SITE $(2V^t/r)$ for a spherical cylinder for different ψ and H

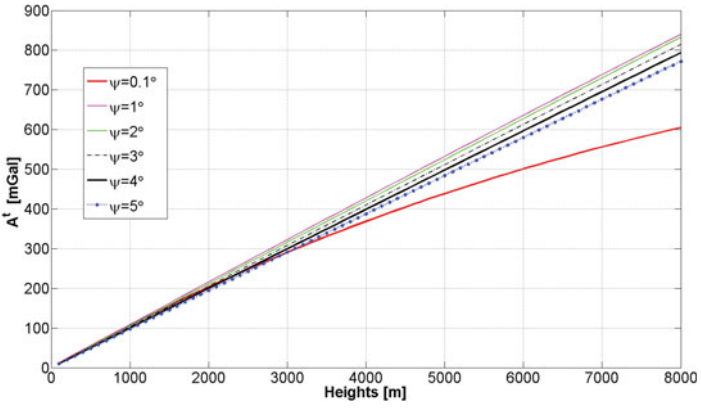

Fig. 8.7 Topographic attraction A^t for different ψ and H

Although the topographic attraction is much larger than the SITE for a small angle ψ, the ratio between SITE and DITE increases with ψ. Also, the spherical contribution of A^t and SITE shows the ratio is:

$$ratio = \left|\frac{SITE}{DITE}\right|_n = \frac{2}{n+1} = \begin{cases} 2 & \text{for} \quad n = 0 \\ 0 & \text{for} \quad n = \infty \end{cases} \tag{8.43}$$

Traditionally, the gravity anomaly can be determined to higher accuracy than the gravity disturbance due to the fact that the latter experienced the uncertainty of the geoid height. This fact is hardly valid nowadays, because: the geoid is rather well-known all over the world (e.g. by EGM2008) and satellite positioning determines directly the geodetic heights of recent gravity data points. As discussed by Sjöberg (2013), the no-topography gravity anomaly is consistent with the Bouguer

gravity disturbance in the sense that both remove all topographic effects. From a practical point of view, it appears that the topographic correction in the Bouguer gravity disturbance is simpler than that for the no-topography gravity anomaly, and consequently Vajda et al. (2007) advocated for using the Bouguer gravity disturbance in gravity inversion. However, as for regional and global studies, the gravity field and topographic heights are usually presented by an EGM and a DEM as series of spherical harmonics, and the workload for determining the disturbance and anomaly are practically the same. Hence, from a practical point of view, one can either work with the isostatic gravity disturbance or the new isostatic gravity anomaly for gravity inversion. In a numerical study of gravity inversion for the Moho determination by the Vening Meinesz-Moritz (VMM) isostatic model (see Sect. 8.5 and Sjöberg 2009), Tenzer and Bagherbandi (2012) demonstrated that the application of the isostatic gravity disturbance improved the RMS fit of the gravimetric solution with the (seismic) CRUST2.0 Moho model compared to the result obtained by using the traditional isostatic gravity anomaly. This numerical result was theoretically explained by Sjöberg (2013) and also Sjöberg et al. (2015). As it is demonstrated here, the improvement in using the gravity disturbance or the refined isostatic gravity anomaly is most significant for gravity inversion in areas with an extended topography, while topographic roughness has less effect, i.e. the effect is most important in the inversion of the low-frequency gravity signal.

8.4 Isostasy

The Earth is a perfect illustration of nature's balance systems, and the principle of isostasy is one example of this. In isostasy, there is a line of equality at which the mass of land above geoid is compensated by the masses below geoid. Therefore, if the Earth is in perfect isostatic balance, there is a depth within the Earth where the total weight of Earth's masses per unit area is the same all around the Earth. This imaginary, mathematical surface is called the depth of compensation (D) below the Earth's surface. A state of isostatic equilibrium in which segments of Earth's crust float at levels determined by their thickness and density is attained by the composition and flow of material in the mantle by the simple formula:

$$\int_{R-D}^{R+H} \rho dh = 0 \tag{8.44}$$

where ρ is the density of the crust and D is the depth of compensation. This simple model suggests that the solid-Earth topography, with height $H > 0$ on the continents and $H < 0$ (as ocean depth) on the ocean, is somehow compensated. As we will see later, the models for compensation varies.

The term isostasy was proposed in 1889 by the American geologist C. Dutton, but the first idea of mass balancing of the Earth's upper layer goes back to Leonardo da

Vinci (1452–1519). The term means that the Earth's topographic mass is balanced (mass conservation) in one way or another, so that at a certain depth the pressure is hydrostatic. Isostasy is an alternative view of Archimedes' principle of hydrostatic equilibrium. According to this principle, a floating body displaces its own weight. A light mountain chain can be compared with an iceberg or a cork floating in water, or, in proper terms, floating on the denser underlying mantle. When a certain area of the crust reaches the "state of isostasy", it is said to be in isostatic equilibrium (balance), and the depth at which isostatic equilibrium prevails is called the compensation depth (Heiskanen and Vening Meinesz 1958, Chap. 5). Although it is generally accepted that the Earth is a dynamic system that responds to loads in many different ways, *isostasy* still provides an important view of the on-going processes.

The Mohorovičić discontinuity, usually called the Moho, is the boundary between the Earth's crust and mantle. This boundary can be determined by isostatic/gravimetric and seismic methods (which do not necessarily fully agree). The masses above the Moho we call the Earth's crust. In 1909, Andrija Mohorovičič, who was a Croatian seismologist, used seismic waves to discover the crust-mantle boundary (the Mohorovičič discontinuity). The Moho separates the oceanic as well as the continental crusts (typically at 10 and 30 km below sea level, respectively) from the underlying mantle. In accurate definition, the Moho is simply a physical/chemical boundary between the crust and mantle where both the crust and mantle are defined by material properties, which can cause large changes in geophysical properties, such as seismic-wave velocity, density, pressure, temperature, etc. (Mooney et al. 1998; Kaban et al. 2003; Martinec 1994).

Usually, topographic loads of wavelengths below 50 km are supported by the underlying lithosphere and are therefore not isostatically compensated. For wavelengths from 50 to 500 km, the topography is typically compensated by elastic flexure in the upper lithosphere, and, for longer wavelengths, it is generally in isostatic balance, except in the very long wavelengths, which are mainly due to dynamic processes in the interior of the mantle.

In geophysics, isostasy is essential mainly for studying geodynamic processes in the crust and upper mantle, and, in geology, it helps in explaining various topographic and geologic features around the world. For example, using isostatic hypotheses, the geological interpretation for compensation of topography and study on structure of lithosphere can be inferred (Ebbing et al. 2012; Kolstrup et al. 2012).

8.4.1 Crustal Thickness and Isostasy

Seismic surveys are expensive, and in many areas seismic information feasible for depth estimation of the crust is therefore sparse or lacking, inferring poor crustal thickness models. This is especially the case in oceanic regions. Today gravity surveys are much more cost-effective, allowing crustal thickness to be estimated by gravity inversion under the assumption of some kind of isostatic model (e.g. the models of Airy, Pratt, Vening Meinesz or Vening Meinesz–Mortiz), where the

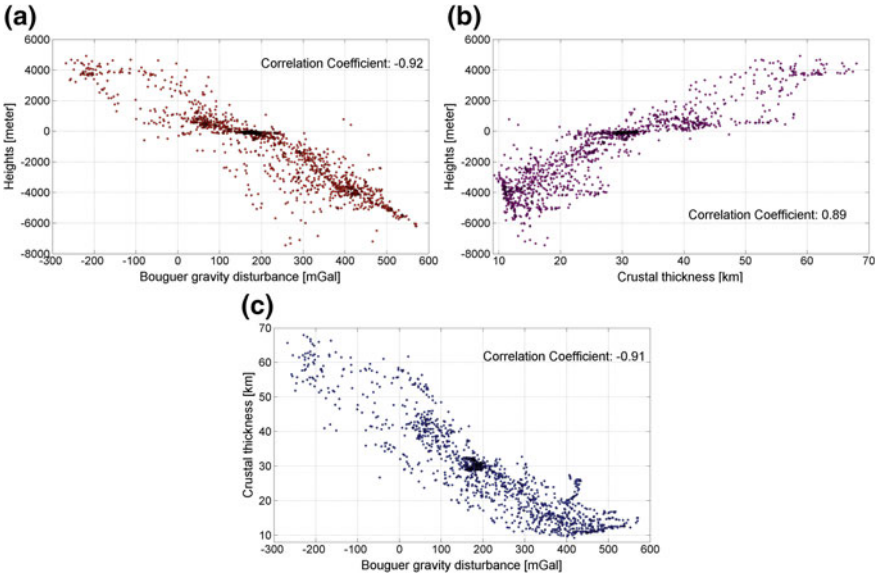

Fig. 8.8 Relations between **a** station elevation, **b** Bouguer gravity disturbance and **c** crustal thickness for seismic stations in South America (based on both land and ocean data)

gravity anomaly is mainly assumed to be the effect of variations in the crustal thickness. The isostatic assumption needs to be verified before using gravity-derived crustal thickness to fill-in data gaps in seismic coverage. Correlations between crustal thickness, height, and Bouguer anomaly are good indicators for the possibility of using gravity-derived crustal thickness, from offshore areas to the highest mountains (Assumpcao et al. 2013b).

Figures 8.8a–c show that the correlations obtained from the compiled datasets in South America follow the expected behaviour for an isostatically compensated crust. Higher elevations correspond to lower Bouguer gravity disturbance and thicker crust (Fig. 8.8a). Considering the large variations between oceanic and continental crust, or between the stable cratonic lowlands and the high South American plateau, verifies that the Bouguer disturbance can be used to determine crustal thickness by applying an isostatic model. The high correlation between the Bouguer gravity disturbance and crustal thickness (the correlation coefficient is more than 0.91 in Fig. 8.8c), shows that the Moho geometry can be estimated by the Bouguer gravity disturbance. Figures 8.9a–c support this idea. (Note that these figures show only the relations between height, Bouguer gravity disturbance and crustal thickness in the continental regions.) Figure 8.9a show a correlation (coefficient) of −0.52 between height and Bouguer gravity disturbance, while Fig. 8.9b indicates a coefficient of −0.54 between Bouguer gravity disturbance and crustal thickness. Hence, the Bouguer gravity disturbance can be a useful indicator of the crustal thickness. On the other hand, Fig. 8.9c shows that there is low correlation (0.35) between the topographic height and crustal thickness, which means that the

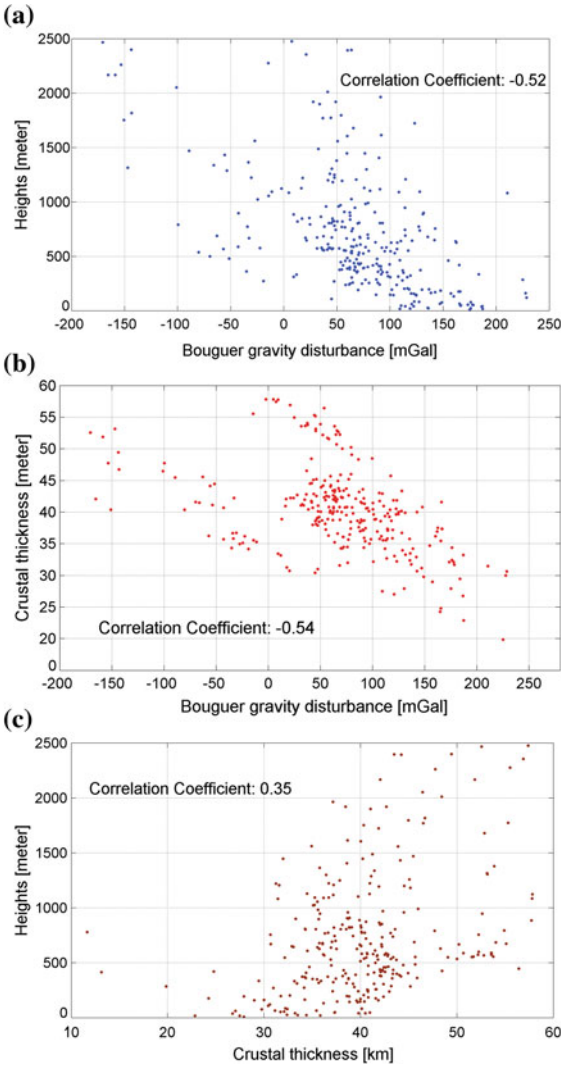

Fig. 8.9 Relations between **a** station elevation, **b** Bouguer gravity disturbance and **c** crustal thickness for seismic stations in South America (based on only continental data)

variations of topography within the stable continental interior are more or less isostatically compensated. Hence, this result puts into doubt the Airy-Heiskanen's hypothesis of a close correlation between topographic height and Moho depth (Sect. 8.4.2). This result also indicates that using local compensation models, such as those of Airy and Pratt, which are only related to topographic height, will frequently not be successful in determining the crustal thickness within continents (at least not in South America). Alternatively, gravity data and an isostatic model could be used for crustal thickness modelling.

8.4.2 Crustal Thickness Models

The original principles of isostasy by Airy (1855) and Pratt (1855) are based on local compensation mechanisms by assuming that the density of a unit prism of the Earth's crust times its volume is constant, i.e. equal-pressure and equal-mass hypotheses at the compensation depth (that varies between the models) (see Eq. 8.44). These models assume that the topographic mass compensation is uniformly distributed vertically and directly compensates the topographic masses along the vertical (local compensation), i.e. the reciprocal forces from the mantle compensate for the pressure of the topographic masses. As the Earth's crust is very complicated, some approximations must be considered for compensating the topographic masses. For example, using a constant density for different layers of the topographic masses is such an approximation. Different hypotheses have been presented based on this principle. The ideal model should be realistic and easy to apply, and it may involve either or both a variable compensation depth or/and a variable crustal density.

There are different methods that can be used to determine the Moho depth by using gravity data. Important references and studies are the solutions of Oldenburg (1974), Shin et al. (2009), Barzaghi and Biagi (2014) at regional scale and the works of Sünkel (1985), Moritz (1990) and Sjöberg (2009, 2013) at global scale. Table 8.2 shows various isostatic and non-isostatic Moho depth models, showing some of their characteristics.

Among the isostatic models, Airy-Heiskanen's and Pratt-Hayford's models are based on classical isostatic assumptions with local compensations (see Figs. 8.15 and 8.16). In contrast, Vening–Meinesz (1931) modified the Airy-Heiskanen theory by introducing a regional instead of local isostatic compensation (see Fig. 8.17). Parker-Oldenburg's method (Parker 1972; Oldenburg 1974) is a gravimetric method that, similar to the VMM method, assumes a variable Moho depth and a constant density contrast. In fact, both models use inversion theory (see, e.g. Dorman and Lewis 1970), implying that the Bouguer gravity anomaly relates to elevation/depth at a certain position. Parker-Oldenburg's method is based on a planar approximation model and uses FFT technique, and it uses an iterative procedure to estimate the Moho depth. The VMM solutions by Moritz and Sjöberg use the Vening Meinesz' idea, but Moritz' solution is an iterative solution while Sjöberg presented a direct solution (see Sect. 8.5). The VMM model is a flexible model, because we can alter the model for estimating either the Moho depth or the density contrast or even both parameters (see Sjöberg and Bagherbandi 2011).

Seismic models are based on measuring the travel-time of the seismic waves (e.g. CRUST1.0). The advantages of using artificial explosions rather than earthquakes are that the time and position of the shot are accurately known (see Fig. 8.10). There are some limitations in the seismic surveys to estimate the Moho depth: (1) The seismic data acquisition is expensive and the amount of data collected in a survey can rapidly become overwhelming; (2) The seismic model suffers from lack of global coverage of data; and (3) The data reduction and processing is time-consuming, requiring sophisticated computer hardware, demanding considerable expertise.

Table 8.2 Comparison of various models for determination of the Moho depth

	Model	Area of compensation	Known parameters	References
Isostatic models	Airy-Heiskanen	Local	Height	Airy (1855)
	Pratt-Hayford	Local	Height and compensation depth	Pratt (1855)
	Vening Meinesz	Regional	Height	Vening Meinesz (1931)
	Sünkel	Global	Modified Airy-Heiskanen	Sünkel (1985)
	Moritz	Global	Gravity and height	Moritz (1990)
	Vening Meinesz-Moritz (KTH method)	Global	Gravity and height	Sjöberg (2009)
Seismic models	Čadek and Martinec	Global	Seismic data	Čadek and Martinec (1991)
	CRUST5.0	Global	Seismic data	Mooney et al. (1998)
	CRUST2.0	Global	Seismic data	Bassin et al. (2000)
	CRUST1.0	Global	Seismic data	Laske et al. (2013)
	European Moho depth	Europa	seismic profiles, 3-D models obtained by body and surface waves	Grad et al. (2009)
	EuCRUST-07	Eurasia	seismic tomography data	Tesauro et al. (2008)
	Antarctica 1×1	Antarctica	seismic data	Baranov and Morelli (2013)
Combined models	KTH-C	Global	Gravity and seismic data	Sjöberg and Bagherbandi (2011)

One of the first global models of the Moho surface was presented by Čadek and Martinec (1991) in terms of the spherical harmonics to degree and order 30 based on different sources of seismic data (see Fig. 8.12).

Because of large data gaps in the seismic models, the isostatic Moho models (e.g. VMM) are complementary to seismic models, in particular in areas where seismic data are sparse. Hence, due to the latest developments in satellite technology and gravimetry, the problem of seismic data gaps can be solved/diminished by gravimetric-isostatic methods. An early combined global Moho model based on seismic and gravimetric data was presented by Sjöberg and Bagherbandi (2011).

However, the gravimetric/isostatic models suffer from the assumed isostatic hypotheses, which are used in formulating the models. This is because it is not clear how realistic the applied hypotheses are, and the real Moho geometry does not necessarily follow those assumptions. However, the rather large differences between the seismic and gravimetric-isostatic methods in some areas are mainly due to sparse seismic data. This is the case in the vast continental regions, where seismic measurements are not available, such as for large portions of Africa, Peru–Chile, Asia and Greenland and also open ocean areas (Mooney et al. 1998). Therefore, in comparison with seismic data, gravimetric methods have the great advantage of having much better access to data at any place on Earth.

8.4.2.1 Seismic Models

As mentioned in the introduction to Sect. 8.4, the boundary between the Earth's mantle and crust is generally approximated by a sharp surface defined as the discontinuity separating rocks having V_p (P-wave) velocities of 6 km/s from those having velocities of about 8 km/s (Meissner 1973). The compositional differences between crust and mantle, causing this increase in seismic velocity, is usually reflected in a corresponding increase of density. Seismic models (e.g. Čadek and Martinec 1991; CRUST5.0; CRUST2.0; CRUST1.0) are based on measuring the travel times of the seismic waves.

8.4.2.2 Seismic Techniques

By seismic methods, it is possible to explore geophysical properties of the Earth's layers and to determine their thicknesses, especially for the crust. Seismic waves can be measured by seismographs at recording stations in both an observatory and in the field. The speed of the wave is calculated and then used to determine the thickness of layers (e.g. Moho depth) and probable composition of the layers. Generally, there are two seismic methods to study the Earth's interior: global seismology (earthquake data) and controlled source seismology (Fig. 8.10). Using earthquake waves provide information on global earth structure and large-scale

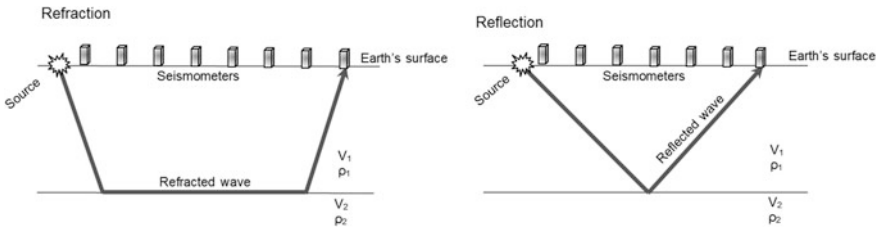

Fig. 8.10 Seismic waves and their behaviour (seismic refraction vs. reflection)

velocity anomalies (100's–1000's km) while using controlled-source seismology enables higher resolution studies (metres to 100's km). The controlled-source seismology requires a controlled seismic source of energy, such as dynamite, a specialized air gun or a seismic vibrator, commonly known by the trademark name Vibroseis. The method uses recorded arrival times at a line or array of geophones to interpret the velocity structure. This method can be divided into two techniques, refraction and reflection techniques which are depended on the seismic waves transmitted through the Earth's layers (see Fig. 8.10):

- The seismic refraction method is based on the measurement of the travel time of seismic waves refracted at the interfaces between subsurface layers of different velocity. Data are recorded on a seismograph and later analysed. Depth profiles for each refractor are produced by an analytical procedure based on consideration of shot and receiver geometry, the measured travel times and calculated velocities.
- The seismic reflection method is based on the fact that when a wave reaches a boundary between two materials having different physical properties (e.g. density or temperature), that wave will be reflected back to the surface. The angle at which the seismic waves are reflected is determined by the angle of incidence of the waves.

There are numerous classes of seismic waves, and they all move in different ways. The two main types of waves are body waves and surface waves (Skinner and Porter 1995). Body waves can travel through the Earth's inner layers, but surface waves only move along the surface of the planet like ripples on water. Earthquakes release seismic energy as both body and surface waves.

The first kind of body wave is the **P** wave or primary wave. It has the greatest velocity of all seismic waves (6 km/s in the uppermost portion of the crust), and, consequently, it is the first signal to be recorded at a seismic station. The **P** wave can move through solid rock and fluids, like water or the liquid layers of the Earth, and it causes alternate compressions and expansions the rock that it passes through. Hence, the particles in the rock will move back and forth parallel to the direction of the wave.

The second type of body wave is the **S** wave (secondary wave or shear wave), which is slower than the **P** wave and can only move through solid rock, not through any liquid. The speed of the wave in the upper crust is about 3.5 km/s. This wave causes a shearing motion, and the particles in a rock will move up and down perpendicular to the direction of the wave propagation.

Variations in seismic wave velocities within the Earth's interior can be seen in Fig. 8.11. Wave velocity changes occur at the boundaries between crust and mantle and between mantle and core. Another change occurs at the depth of about 100 km (corresponding to the lithosphere–asthenosphere boundary) and at the 400 and 670 km deep discontinuities. For example, the P-wave velocity at the top of the mantle is about 8 km/s, and it increases to 14 km/s at the core–mantle boundary.

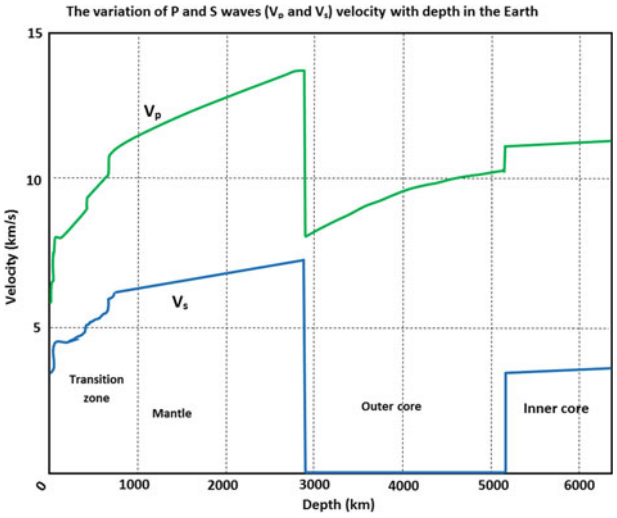

Fig. 8.11 Seismic wave velocities in the Earth (Kennett and Engdahl 1991). Variations of P- and S-wave velocities with depth

8.4.2.3 The Crustal Thickness Model of Čadek and Martinec

A Moho model, compiled from various seismic data sources, is shown in Fig. 8.12. The source material used for constructing the spherical harmonic expansion of the crustal thickness was the seismic data produced by Meissner et al. (1987), Belyaevsky (1981), Belyaevsky and Volkovsky (1980), Allenby and Schnetzler (1983) and Goslin et al. (1972). In no data regions a mean value of the crustal thickness was assumed. For example, they assumed a uniform crustal thickness of 7 km in the Arctic Ocean, Southern Ocean, Indian Ocean, South Atlantic Ocean and 35 km in North of Canada (see more details in Čadek and Martinec 1991, Fig. 1). For oceanic regions with a considerable local topography (i.e. for islands), an appropriate topographic correction was considered (e.g. in New Zealand). The model was presented as a global crustal-thickness map in a grid with steps sized 2°.

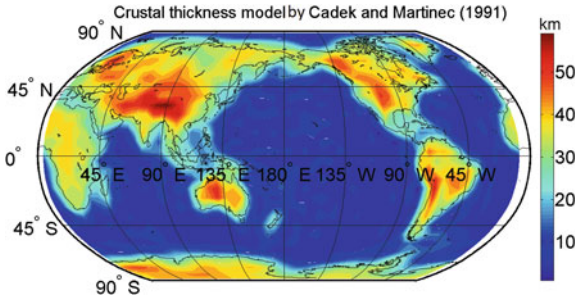

Fig. 8.12 Crustal thickness model by Čadek and Martinec (1991) in terms of the spherical harmonics to degree and order 30 based, on various sources of seismic data with a grid size of 6° × 6°. Unit: km

As the quality of the seismic data is not uniform and the observation error affects the result, they presented a simple method to find an optimal upper degree $n_{\max}$. First they introduced the Root Mean Square difference (Δ) between the crustal depth data sets $D(\theta_i, \lambda_i)$ and the model given by the finite spherical harmonic expansion:

$$\Delta = \left[\frac{1}{N} \sum_{i=1}^{N} \left[D(\theta_i, \lambda_i) - \sum_{n=0}^{n_{\max}} \sum_{m=-n}^{n} d_{nm} Y_{nm}(\theta_i, \lambda_i) \right]^2 \right]^{1/2}, \tag{8.45}$$

where N is the number of observation points and $n_{\max}$ is the maximum degree of the model.

The data set consists of observations distributed globally. Using trial and error and changing the degree and order, they varied Δ. The obtained Δ should be compared with a threshold value to estimate $n_{\max}$. They assumed the average error Δ_O = 2–3 km on the oceans and Δ_C = 5 km on the continents. Accepting these error estimates, they found $n_{\max}$ to be about 30. In this case, $\Delta \leq \min(\Delta_C, \Delta_O)$. The coefficients d_{nm} are available by following the above mentioned method for $n_{\max}$ = 30. The numerical results are those illustrated in Fig. 8.12.

8.4.2.4 CRUST5.0

Another global seismic model is CRUST5.0 (Mooney et al. 1998), including 2592 tiles at a resolution of 5° × 5°, was created as a first attempt to estimate the global crustal thickness (see Fig. 8.13). CRUST5.0 is based on seismic refraction data published in the period 1984–1995. The accuracy of CRUST5.0 is not specified, but it varies from place to place. For example, the accuracies of this model seems to be generally better in the United States and Europe because of denser seismic measurements (Mooney et al. 1998, Fig. 1), but much worse in Africa, Greenland and in some parts of Asia and Antarctica. For the latter regions, the crustal thickness was only estimated by some interpolation method.

8.4.2.5 CRUST2.0

The global crustal model CRUST2.0 was released by the US Geological Survey and Institute for Geophysics at the University of California in 2000 (Bassin et al. 2000; Laske et al. 2000). It is a global crustal model with a grid of 2° × 2°, offering a rather detailed density structure of the crust and uppermost mantle. CRUST2.0 is obtained by the analysis of the travel times of seismic waves. This model takes advantage of the compilation of the global sediment thickness, which is defined on a 1° × 1° grid as well as compiled ice thickness on the same scale. This model includes information on the depth and density of layers: (1) ice, (2) water, (3) soft

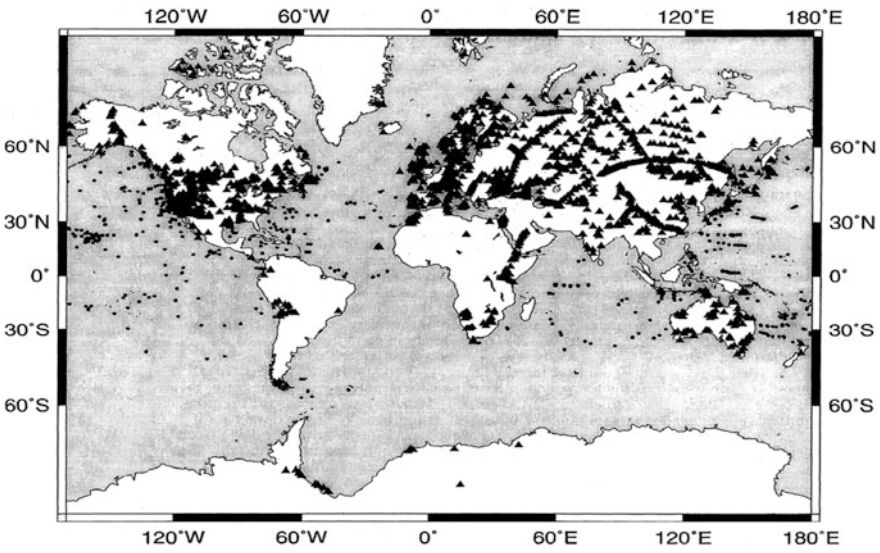

Fig. 8.13 CRUST 5.0 data, location of seismic refraction profiles used in Mooney et al. (1998). Triangles correspond to locations within continents and on margins. *Solid circles* are locations of oceanic refraction profiles (after Christensen 1982)

sediments, (4) hard sediments, (5) upper crust, (6) middle crust and (7) lower crust, whose 7th layer describes the Moho depth. The model does not cover the entire Earth (Mooney et al. 1998), and it was estimated by interpolation method.

8.4.2.6 CRUST1.0

The Moho depth in CRUST1.0 is based on one-degree averages of crustal thickness data from active-source seismic studies (earthquake data) as well as from controlled–source seismology. In areas where such constraints are still missing, for example in Antarctica, the crustal thicknesses are estimated using gravity constraints. In CRUST1.0, the principal crustal types adopted from both CRUST2.0 and CRUST5.0, as well as from the crustal types aligned with the crustal-age map of Artemieva and Mooney (2001). CRUST1.0 also has a few new types of data, including one for very young (<3 Myears) oceans. In contrast to older models, the function of crustal types in CRUST1.0 is limited to assigning elastic parameters to layers in the crystalline crust. There are less than 40 crustal types, and each of the $1^\circ \times 1^\circ$ cells have a unique eight-layer crustal profile consisting of the layers for water, ice, upper sediments, middle sediments, lower sediments, upper crust, middle crust and lower crust.

The mean Moho depths of CRUST1.0 are 23 km (global), 38.0 km (in continents) and 12.6 km (in oceans), and the minimum and maximum crustal thicknesses are 7.4 and 75 km, respectively (see Fig. 8.14).

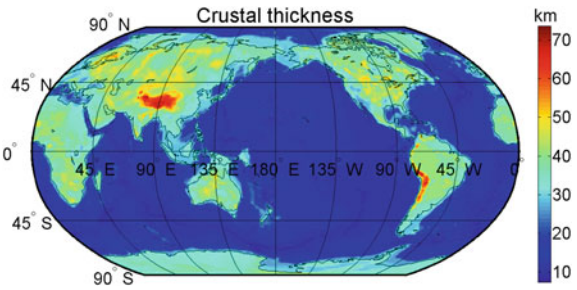

Fig. 8.14 CRUST1, global crustal model at 1° × 1° (Laske et al. 2013)

8.4.2.7 Airy-Heiskanen's Hypothesis

According to Airy-Heiskanen's hypothesis (Heiskanen and Moritz 1967, Sect. 3.4), the solid Earth's crust floats on the viscous and denser mantle exactly as icebergs in the oceans. This hypothesis puts emphasis on local compensation, and there are light roots under mountains and heavy anti-roots under oceans (Fig. 8.15). Although both models originate from Eq. (8.44), the fundamental difference between Airy-Heiskanen's and Pratt-Hayford's views is that the former postulates a uniform density with varying thickness of crust, while the latter is constructed by a uniform depth with varying crustal density.

Considering Eq. (8.44) for density ρ_c in the crust and $\Delta\rho = \rho_m - \rho_c$, that is the density contrast between mantle (ρ_m) and crust (the density of the mantle 3.27 g/cm^3 is 18% denser than that of the crust), we obtain the following equation below compensation depth D_0:

$$H\rho_c - t_l\Delta\rho = 0 \tag{8.46a}$$

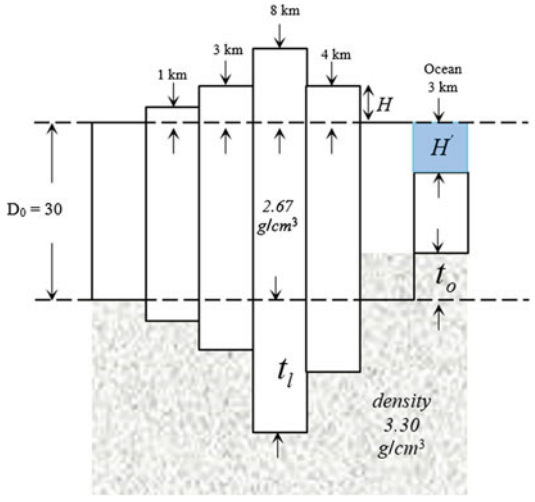

Fig. 8.15 Airy's model. The solid-Earth topographic features are compensated by roots (t_l) over land and anti-roots (t_O) over ocean areas

which yields the root

$$t_l = \frac{H}{\Delta\rho}\rho_c, \tag{8.46b}$$

where H is the topographic height, and t_l is root of the crust. In the Airy-Heiskanen's hypothesis, the density of the crust is assumed constant.

At sea, the equilibrium condition (8.44) can now be written as:

$$d(\rho_c - \rho_w) + t_o\Delta\rho = 0 \tag{8.47a}$$

where d is sea depth and ρ_w is the density of water, thus the anti-root (t_o) is given by:

$$t_o = \frac{d(\rho_c - \rho_w)}{\Delta\rho}, \tag{8.47b}$$

The Earth's crustal thickness D will vary by the thicknesses of roots and anti-roots (Heiskanen and Vening Meinesz 1958, p. 137):

$$D = \begin{cases} D_0 + t_l + H & \text{continental} \\ D_0 - t_o - H' & \text{ocean} \end{cases}, \tag{8.48}$$

where D_0 is the normal Moho depth and H' denotes the depth of the ocean. The roots and anti-roots of the crustal thickness with respect to D_0 can also be obtained in a spherical approximation (Rummel et al. 1988):

$$t_l = \rho^* H\left\{1 + \frac{2D_0 + (\rho^* + 1)H}{R} + \frac{(2D_0 + \rho^* H)[2D_0 + (\rho^* + 1)H]}{R^2} - \frac{D_0(D_0 + \rho^* H)}{R^2} - \frac{(\rho^{*2} - 1)H^2}{3R^2}\right\}, \tag{8.49a}$$

$$t_o = \rho' H'\left\{1 + \frac{2D_0 - (\rho' + 1)H'}{R} + \frac{(2D_0 - \rho' H')[2D_0 + (\rho' + 1)H']}{R^2} - \frac{D_0(D_0 - \rho' H')}{R^2} - \frac{(\rho'^2 - 1)H'^2}{3R^2}\right\}, \tag{8.49b}$$

where $\rho^* = \rho/\Delta\rho$ and $\rho' = (\rho_c - \rho_w)/\Delta\rho$, and R is the Earth's mean radius. The planar approximations in Eqs. (8.46b) and (8.47b) are obtained by letting R approach infinity in these equations.

8.4.2.8 Pratt-Hayford's Model

According to Pratt-Hayford's hypothesis, the isostatic compensation depth (D) is constant (typically 100 km) and the density (ρ) in and under the mountains is smaller than under the flat regions (ρ_0) or under the oceans (Fig. 8.16). Here the density variations play a main role in the isostatic system. Also, the compensation layer is located directly underneath the mountains and reaches to the compensation depth, where equilibrium prevails (Heiskanen and Vening Meinesz 1958, p. 126).

Mathematically, the equilibrium condition Eq. (8.44) can be expressed in continental and ocean regions as:

$$(D+H)\rho = D\rho_0 \Rightarrow \rho = \frac{D}{D+H}\rho_0 \tag{8.50a}$$

and

$$(D-H')\rho + H'\rho = D\rho_0 \Rightarrow \rho = \rho_0 + \frac{H'}{D-H'}(\rho_0 - \rho_w), \tag{8.50b}$$

respectively.

In conclusion, the classical isostatic models of Airy and Pratt for the topographic masses are compensated either by a variable Moho depth or crustal density. Pratt-Hayford's model is based on a constant compensation depth and a variable topographic density (or density contrast), while Airy-Heiskanen's model assumes a constant topographic density with a variable compensation depth (mountain root). In reality, both Moho depth and density contrast vary. The best model is the one that considers both the Moho depth and density contrast as variables. A disadvantage of the original Airy-Heiskanen model is the fact that the anti-roots in the ocean areas can attain unrealistic values that are larger than the thicknesses of the crust (Makhloof 2007). This means that H' must not be larger than approximately 8000 m.

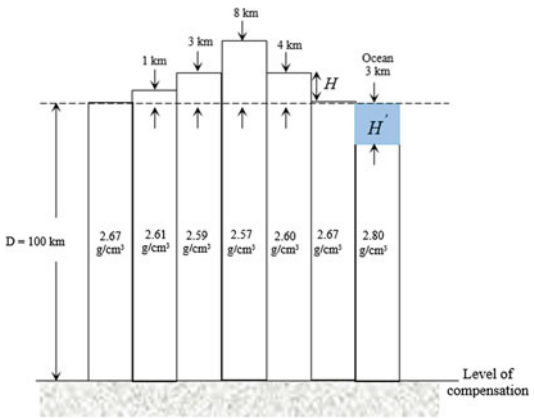

Fig. 8.16 Pratt's model. The solid–Earth-crust elevation variables (H and H') are compensated by a variable density above the compensation depth D

This problem primarily occurs in some deep-ocean areas; for example, in the Arctic Ocean with 5450 m depth, the 7725 m Java trench in the Indian Ocean, the 8648 m Puerto Rico trench in the Atlantic Ocean and the 11033 m Mariana trench in the Pacific Ocean. Usually, a combined Airy-Pratt model is applied to overcome such a problem (Wild and Heck 2004a, b; Makhloof 2007) by implementing the Airy-Heiskanen model for the continental region and Pratt-Hayford model for the ocean areas. Also, Airy-Heiskanen's compensation model usually assumes a constant density contrast at the Moho discontinuity, although a laterally variable density contrast should be more realistic.

8.4.2.9 Vening Meinesz' Model

Vening Meinesz (1931) modified the Airy-Heiskanen theory by introducing a regional isostatic compensation based on a flat-Earth approximation. In his approach, each topographic column of an infinitely small cross-section is considered to be compensated by a mass equal to the local compensation of this element but spread horizontally according to the curved deformation of the Earth's crust. Each inverse problem in isostasy assumes the isostatic gravity anomaly to be zero under a certain isostatic hypothesis. In the case of the Vening Meinesz isostatic hypothesis, the Moho *density contrast is constant*, while the *Moho depth is variable* (in Airy's model). Hence, the Vening Meinesz inverse isostatic problem aims at determining a suitable variable Moho depth for a prescribed constant-density contrast. Seismology shows that the crust is thicker (30–90 km) underneath continents than under oceans (5–15 km). This variation of the depth of the Moho and the general agreement between the thickness of the crust estimated from the seismic and gravity methods support the Airy-Heiskanen and Vening Meinesz models. In the Airy-Heiskanen theory, there is no correlation between neighbouring crustal columns, while we know that this must be the case in reality due to the elasticity of the Earth. The difference between the Airy-Heiskanen and the Vening Meinesz models is thus a matter of local-versus-regional mechanisms of topographic mass compensation. In the regional compensation model, the compensating masses of the mountains are distributed laterally, and, to achieve this, Vening Meinesz (1931) assumed that the crust is a homogenous elastic plate floating on a viscous mantle (Fig. 8.17).

In Vening Meinesz' model, local compensation spreads horizontally, reducing from the centre of the region towards the margin. Therefore, bending of the crust is expected for an elastic plate floating on the mantle (Vening Meinesz and Heiskanen 1958, pp. 138–139). The main assumption in the Vening Meinesz model is that the load of the topographic mass causes the Earth's crust to bend until equilibrium prevails. The bending curve can be determined from Hertz's formula by assuming that the plate of infinite dimensions is loaded by a concentrated mass. This model shows how density compensates with the distance from the load centre (Fig. 8.18). The sinking (S) in the centre of the load is given by:

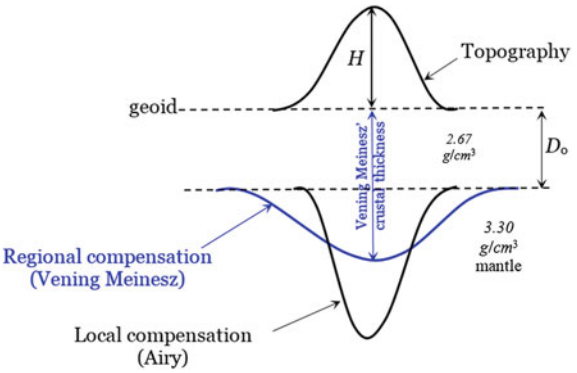

Fig. 8.17 Local and regional compensation corresponding to Airy and Vening Meinesz hypotheses

$$S = \frac{\rho_c}{8(\rho_m - \rho_c)L^2} \tag{8.51}$$

where L is the load's width. The sinking will decrease with distance from the centre. Figure 8.18 shows the relation between increasing the distance from the centre (x), where x is function of L, and sinking factor n (where n = 1, 2, …, m).

8.4.2.10 Sünkel's Model

The difference between the Airy-Heiskanen and the Vening Meinesz models is a matter of local-versus-regional mechanisms of topographic compensation. Already some studies have been performed for modifying Airy-Heiskanen's model. The most well-known is Sünkel's (1985) model, which can be also found in Rummel et al. (1988). Sünkel estimated the Earth crustal thickness by introducing a smoothing factor in Airy-Heiskanen's model to obtain a regional compensation according to Vening Meinesz' idea. Rummel et al. (ibid) studied global topographic/isostatic models and defined some criteria (e.g. that the norm of the

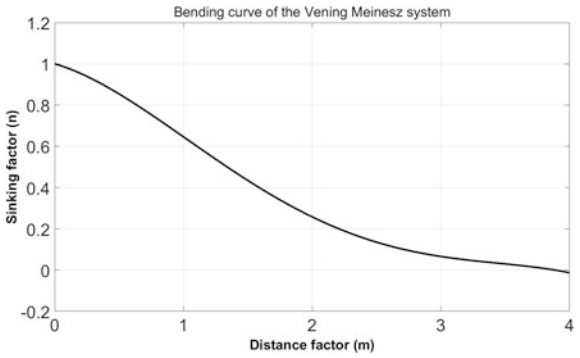

Fig. 8.18 The sinking at the distance x from the centre (e.g. $x = 1.0L, 2.0L, \ldots, mL$)

residual potential of topographic and compensating masses should be a minimum) to estimate the depth of isostatic compensation.

According to Sünkel (1985), the topographic-isostatic (TI) harmonic coefficients based on the Airy-Heiskanen model are given by (Rummel et al. 1988, see also Sjöberg 1998a, b):

$$c_{nm}^{TI} = \frac{3}{2n+1}\frac{\rho_c}{\rho_e}\left[\left(1-\left(\frac{R-D}{R}\right)^n\right)\frac{H_{nm}}{R} + \frac{n+2}{2}\left(1+\frac{\rho_c}{\Delta\rho}\left(\frac{R-D}{R}\right)^{n-3}\right)\frac{(H^2)_{nm}}{R^2} + \frac{(n+2)(n+1)}{6}\left(1-\left(\frac{\rho_c}{\Delta\rho}\right)^2\left(\frac{R-D}{R}\right)^{n-6}\right)\frac{(H^3)_{nm}}{R^3}\right]. \tag{8.52}$$

where D is the crustal thickness (or depth of compensation).

By assuming that the Vening Meinesz (VM) model is correct, the smoothed Airy-Heiskanen (SAH) model can be obtained by assuming the following harmonic coefficients for AH and VM

$$c_{nm}^{\mathrm{AH}} \approx \frac{3}{(2n+1)\rho_{\mathrm{e}}}\left[\frac{(\bar{\rho}_{\mathrm{c}}H)_{nm}}{R} + \frac{(n+2)D_0}{R^2}(\bar{\rho}_{\mathrm{c}}H)_{nm} + \frac{(n+2)}{2R^2}\frac{1}{\rho_{\mathrm{m}}-\rho_{\mathrm{c}}}\left(\bar{\rho}_{\mathrm{c}}^2H^2\right)_{nm}\right] \tag{8.53a}$$

$$\mathrm{c}_{nm}^{VM} \approx \frac{3}{(2n+1)\rho_e}\left[\frac{(\Delta\rho(D_{VM}-D_0))_{nm}}{R} + \frac{(n+2)}{2}\frac{\left(\Delta\rho(D_0^2-D_{VM}^2)\right)_{nm}}{R^2} + \frac{(n+2)(n+1)}{6}\frac{\left(\Delta\rho(D_{VM}^3-D_0^3)\right)_{nm}}{R^3}\right], \tag{8.53b}$$

the SAH harmonic coefficients (c_{nm}^{SAH}) according to Sünkel (1985, p. 23) can be written as:

$$c_{nm}^{SAH} = \frac{3}{2n+1}\frac{\rho_c}{\rho_e}\left[\left(1-\beta_n\left(\frac{R-D}{R}\right)^n\right)\frac{H_{nm}}{R} + \frac{n+2}{2}\left(1+\beta_n\frac{\rho_c}{\Delta\rho}\left(\frac{R-D}{R}\right)^{n-3}\right)\frac{(H^2)_{nm}}{R^2} + \frac{(n+2)(n+1)}{6}\left(1-\beta_n\left(\frac{\rho_c}{\Delta\rho}\right)^2\left(\frac{R-D}{R}\right)^{n-6}\right)\frac{(H^3)_{nm}}{R^3}\right]. \tag{8.54}$$

where $GM = 4GR^3\rho_e\pi/3$ is the geocentric gravitational constant, $\bar{\rho}_c$ is the crust density, $\rho_e \approx 5.5$ g/cm^3 is the mean density of the Earth's mass and β_n is the degree-dependent smoothing factor. This corresponds to an isotropic smoothing operator

$$B(t) = \frac{1}{4\pi}\sum_{n=0}^{\infty}(2n+1)\beta_n P_n(t), \tag{8.55a}$$

with eigenvalues β_n:

$$\beta_n = 2\pi \int_{-1}^{1} B(t)P_n(t)dt. \quad t = \cos\psi \tag{8.55b}$$

The operator B can be used in both spatial and spectral form:

$$B(t)H \leftrightarrow \beta_n H_{nm}. \tag{8.55c}$$

Here we should notice that the rock/ocean topography remains unchanged, and only the compensation root/antiroot surface is smoothed. The idea behind the operator B is to replace the block-averaging operator by a homogenous and isotropic moving-average operator of constant weight. According to the convolution theorem, the moving-average convolution process in the space domain corresponds to a simple product between the eigenvalues of the moving-average operator and the Fourier coefficients of the function to be averaged in the frequency domain (Sünkel 1985).

8.4.2.11 An Alternative Method Using Sünkel's Model

For estimating the smoothing factors, we form the following equation based on the Airy-Heiskanen and the VMM models (Sect. 8.5)

$$c_{nm}^{VMM} - c_{nm}^{SAH} = 0, \tag{8.56a}$$

where c_{nm}^{VMM} and c_{nm}^{SAH} are the VMM and the smoothed AH harmonic coefficients, respectively.

Considering only the two first terms in Eq. (8.54), the smoothed isostatic coefficients can be written as:

$$\left(c_{nm}^{si}\right)_{AH} = \frac{3}{2n+1}\frac{\rho_c}{\rho_e}\left[\beta_{nm}\left(\frac{R-D}{R}\right)^n \frac{H_{nm}}{R} - \beta_{nm}\frac{n+2}{2}\frac{\rho_c}{\Delta\rho}\left(\frac{R-D}{R}\right)^{n-3}\frac{(H^2)_{nm}}{R^2}\right], \tag{8.56b}$$

where β_{nm} is the smoothing factor for each degree and order which can be can be uniquely solved from this equation. By inserting Eq. (8.56b) into Eq. (8.56a), the smoothing factors can be obtained using the least-square adjustment for each degree:

$$\hat{\beta}_n = \sum_{m=-n}^{n} \frac{(2n+1)\rho_e c_{nm}^{VMM}}{3\rho_c\left(\alpha_{nm}^1 - \alpha_{nm}^2\right)}, \tag{8.57a}$$

where:

$$\alpha_{nm}^{1} = \left(\frac{R-D}{R}\right)^{n} \frac{H_{nm}}{R}, \tag{8.57b}$$

$$\alpha_{nm}^{2} = \frac{n+2}{2}\frac{\rho_c}{\Delta\rho}\left(\frac{R-D}{R}\right)^{n-3} \frac{H_{nm}^{2}}{R^{2}}, \tag{8.57c}$$

Thus the harmonic coefficients of the VMM can be written by:

$$c_{nm}^{VMM} = \frac{3}{2n+1}\frac{\rho_c}{\rho_e}\hat{\beta}_n\left(\alpha_{nm}^{1} - \alpha_{nm}^{2}\right). \tag{8.57d}$$

The variance of the harmonic coefficients of the compensating potential using Eq. (8.53b) is obtained by:

$$\left(\sigma^{i}\right)_{\mathrm{nm}}^{2} \approx \left(\frac{3}{2n+1}\frac{\Delta\rho}{\rho_{\mathrm{e}}}\right)^{2}\left(\frac{\sigma_{\Delta D_{nm}}^{2}}{R^{2}}\right), \tag{8.58}$$

where $\Delta D = \mathrm{D} - \mathrm{D}_0$ and $\sigma_{\Delta D_{nm}}$ is the standard error of ΔD_{nm}. The harmonic coefficients of ΔD are determined by:

$$(\mathrm{D}-\mathrm{D}_0)_{nm} = \Delta D_{nm} = \frac{1}{4\pi}\iint_{\sigma}(\mathrm{D}-\mathrm{D}_0)Y_{nm}(P)d\sigma = \left(\mathbf{b}_{\mathrm{nm}}^{\mathrm{Y}}\right)^{*}\Delta\mathbf{D}, \tag{8.59}$$

where $\Delta\mathbf{D} = \mathbf{D} - \mathbf{D}_0$. The integral of Y_{nm} in the equation has been vectorized by $\mathbf{b}_{\mathrm{nm}}^{\mathrm{Y}}$, which is multiplied by Moho topography $\Delta\mathbf{D}$ and $()^{*}$ denotes a transpose operator. In this case, the dimension of $\mathbf{b}_{\mathrm{nm}}^{\mathrm{Y}}$ is $2M^2 \times 1$, where M is the number of grid cells from north to south, which is determined according to the resolution. Similarly, the harmonic-coefficient standard error of ΔD_{nm} is estimated by using the law of error propagation:

$$\sigma_{\Delta D_{nm}}^{2} = \left(\mathbf{b}_{nm}^{Y}\right)^{*}\mathbf{Q}_{\Delta D_{nm}}\mathbf{b}_{nm}^{Y}, \tag{8.60}$$

where $\mathbf{Q}_{\Delta D}$ is variance–covariance matrix of ΔD_{nm}.

8.4.2.12 A Practical Method

A practical method was also suggested by Sünkel (1985) for determining β_n through the following condition, which uses the least-squares technique. In this method, the energy of the gravity field reduced by the TI (topographic-isostatic)

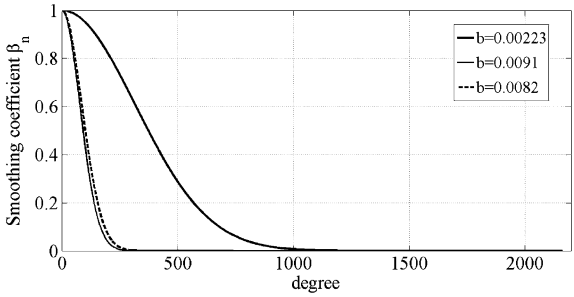

Fig. 8.19 The smoothing factor β_n of Eq. (8.63) estimated by various smoothing parameters b

potential is minimized. Then, by the following formula, the best β_n for estimating the SAH model is sought:

$$\left\| V - V^{SAH} \right\|^2 = \min_D, \tag{8.61}$$

where V^{SAH} and V are the SAH and Earth's gravitational potentials, respectively. According to Sünkel the low-order harmonics of the anomalous gravitational potential are, to a large degree, due to density disturbance in the upper mantle, and probably due to even deeper sources. Therefore, it makes no sense to include only the long-wavelength part in Eq. (8.61) and requiring that $c_n \approx c_n^{SAH}$ and yielding β_n as (based on correlation consideration between the two degree variances):

$$\beta_n = \left(\frac{R - D}{R} \right)^{-n} \left(1 - \frac{\rho_e}{3\rho_0} (2n+1) \frac{c_n^2}{H_n} \right), \tag{8.62}$$

where c_n^2 and $\left(c_n^{SAH}\right)^2$ are the degree variances of Earth's gravitational potential and the SAH model. Alternatively, one can model β_n by a Gaussian filter (Sünkel 1985, p. 23)

$$\beta_n = \mathrm{e}^{-b^2 n^2}, \tag{8.63}$$

where a solution to b (= 0.00223 for D = 24 km) was presented by Sünkel (1985), based on a least squares method. Figure 8.19 shows the smoothing factor as a function of degree n. For larger numbers of b, it decreases faster to zero. In addition, it shows that most power of the TI harmonic coefficients up to value 0.5 will remain, and, below 0.5, it is filtered.

8.5 Moho Determination by Vening Meinesz-Moritz Theory

Moritz (1990, Sect. 8) generalized Vening Meinesz' isostatic hypothesis and solution in a global, spherical approximation, here called Vening Meinesz-Moritz' (VMM) model. The VMM isostatic (flexural) model represents a more realistic

assumption of the global compensation mechanism described for the Earth's homogenous crust. The main idea is simple, but the theoretical modelling is somewhat difficult, because the mass distribution of the Earth's crust is complicated, and also many geophysical phenomena should be considered. Sjöberg (2009) formulated this problem as that of solving a non-linear Fredholm integral, also an equation of the first kind, and he presented some solutions for the crustal thickness and then a solution for the Moho density contrast (MDC) that were published by Sjöberg and Bagherbandi (2011). Tenzer and Bagherbandi (2012) and Sjöberg (2013) suggested Bouguer gravity disturbance rather than Bouguer gravity anomaly as the primary observable in the solutions (see Sect. 8.3), and this idea will be followed here. The VMM problem and its solution are described next.

8.5.1 Formulating the Mathematical Problem

The VMM problem (Moritz 1990) is to determine the Moho depth D such that the compensating attraction A_C totally compensates the Bouguer gravity disturbance δg^B on the Earth's surface (approximated by a sphere of radius R), implying that the isostatic gravity disturbance (δg^I) vanishes for each point on the Earth's surface, i.e.

$$\delta g^I = \delta g^B + A_C = 0. \tag{8.64a}$$

Actually, this equation can be generalized to any point (P) in space with radius equal to or larger than sea-level radius for the isostatic gravity disturbance, gravity anomaly or disturbing potential:

$$\delta g^I(P) = 0; \quad r_P \geq R \tag{8.64b}$$

$$\Delta g^{I,ref}(P) = 0; \quad r_P \geq R \tag{8.64c}$$

and

$$T^I(P) = 0; \quad r_P \geq R \tag{8.64d}$$

These equations are only theoretically exact, but not in reality. First, it is assumed that the δg^B can been reduced for all topographical attraction (A^T). Second, local deviations from isostatic equilibrium may occur.

8.5.2 Formulating the Integral Equation

The isostatic inverse problem according to the VMM hypothesis is primarily based on a constant density contrast ($\Delta\rho = \rho_m - \rho_c = const$) at the Moho boundary, and

a variable Moho depth (Moritz 1990). The integral equations corresponding to Eqs. (8.64b, d) and the radial gravity gradient are all given by Eq. (8.24). For the Bouguer gravity disturbance observable on the sphere of radius R, it becomes:

$$R \iint_{\sigma} K(\psi, \tau) d\sigma = h(P), \tag{8.65a}$$

where:

$$h(P) = -\left(\delta g^{B}(P) + A_{C0}(P)\right)/(G\Delta\rho) \tag{8.65b}$$

$$K(\psi, \tau) = F(R, \psi, D) = \sum_{n=0}^{\infty} \frac{n+1}{n+3}\left[1 - (1-\tau)^{n+3}\right] P_n(\cos\psi);$$
$$s = 1 - \tau; \tau = D/R, \tag{8.65c}$$

or, in a closed form with $t = \cos\psi$ (Sjöberg 2009):

$$K(\psi, \tau) = J_2(t, \tau) - J_3(t, \tau) \tag{8.65d}$$

with

$$J_2(t, \tau) = R^2\left[\frac{(1-2t^2)(1-\tau)+t}{(1-t^2)l_{Px_1}} - \frac{(1-2t^2)+t}{(1-t^2)l_{P0}}\right] + R\ln\Psi, \tag{8.66a}$$

and

$$J_3(t, \tau) = \frac{R^2 t}{1-t^2}\left[\left\{(x/R)^2(1-t^2) - (5-6t^2)\frac{x}{R}t + 2 - 3t^2\right\}\frac{1}{l_{Px}}\right]_{x=x_1}^{x=R} + 3Rt^2\ln\Psi, \tag{8.66b}$$

where:

$$\Psi = \frac{1 - t + l_{P0}/R}{1 - t - \tau + l_{Px_1}/R}. \tag{8.66c}$$

In a similar way, Eqs. (8.21) and (8.24) can be used to formulate integral equations for the Moho depth with Bouguer disturbing potential and radial disturbing anomaly and its gravity gradient as observables.

8.5.3 Solving the Integral Equation to Second Order

If the gravity field is represented by an EGM, the solution for the VMM problem of the Moho depth is conveniently performed in the spectral domain according to the basic derivation in Sect. 8.2 with the solution given, e.g. by Eqs. (8.11b) and (8.12), which we will use to second order (see also Sjöberg 2009, 2013). Here we assume that the Bouguer gravity disturbances in Eq. (8.65b) are available as a series in spherical harmonics, so:

$$h(P) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm} Y_{nm}(P). \tag{8.67}$$

Then a second-order solution for the Moho depth can be written (Sjöberg 2009, 2013) as:

$$D_P = (D_1)_P + \frac{1}{R} \sum_{n=0}^{\infty} \frac{n+2}{2} \sum_{m=-n}^{n} \left(D_1^2\right)_{nm} Y_{nm}(P), \tag{8.68a}$$

or

$$D_P = (D_1)_P + \frac{(D_1)_P^2}{R} - \frac{1}{32\pi R} \iint_{\sigma} \left[\frac{(D_1)^2 - (D_1)_P^2}{\sin^3(\psi/2)} \right] d\sigma, \tag{8.68b}$$

where:

$$D_1 = \frac{1}{4\pi} \sum_{n=0}^{\infty} \frac{2n+1}{n+1} \sum_{m=-n}^{n} h_{nm} Y_{nm}(P) \tag{8.69}$$

is the first-order solution.

The gravimetric/isostatic-induced Moho depth is a smooth surface, whose harmonics beyond, say, degree and order 100, are not very significant (see, e.g. Turcotte and Schubert 2002, p. 195). Therefore the infinite sum should be limited, say, to degree 100–180 (the upper limit is not clarified). Higher-degree terms are likely to be isostatically uncompensated, so that their inclusion in the above equations will lead to false short-wavelength Moho geometry.

A first order approximation (D_1) of the Moho depth can be directly computed from a global set of spherical harmonics of the no-topography gravity disturbance, h_{nm}, as the first (linear) term of Eq. (8.68a, 8.68b), but, practically, by using the equation with a finite upper limit. In practice, the free-air gravity disturbance (δg) is directly available from an Earth gravitational model (EGM), e.g. EGM2008 complete to degree and order 2160 (Pavlis et al. 2012). With access to a corresponding

spherical-harmonic representation H_{nm} of the topographic height (such as DTM2006; Pavlis et al. 2007), the no-topography harmonic coefficients can be computed from the free-air gravity-disturbance coefficients (G_{nm}). (A better solution requires that the lateral density variations of topographic density be considered.) Thus one obtains:

$$h_{nm} = \left[2\pi(\mu H)_{nm} - G_{nm}\right]/(G\Delta\rho). \tag{8.70}$$

As the first/linear term in Eq. (8.69), D_1 is much larger than the remaining second-order terms, the second-order solution is directly obtained (without iteration).

Finally, we present also Eq. (8.69) in the space domain (for $n_2 \to \infty$) as:

$$D_1 = \frac{h(P)}{2\pi} - \frac{1}{(4\pi)^2} \iint_{\sigma} H(\psi)h(Q)d\sigma, \tag{8.71}$$

where the second term is a Hotine type integral (Sect. 3.3.3; Hotine 1969, p. 392).

In Eq. (8.68b), the integral should be significant only in a small region around the computation point. As Eq. (8.68b) includes all terms of second order and $D \leq$ 100 km, the achievable accuracy due to this approximation is of the order of $100^3/6371^2$ km $\approx$ 25 m, which should be sufficient for most practical applications. The third term Eq. (8.68b) needs some special care to compute, as it has a strong singularity at the computation point. Hence, for the near-zone (within the cap size ψ_0), we may use a plane approximation to the integral (I), i.e.

$$I = \frac{1}{32\pi R} \int_{\alpha=0}^{2\pi} \int_{\psi=0}^{\psi_0} \frac{D^2 - D_P^2}{\sin^3 \frac{\psi}{2}} \sin\psi d\psi d\alpha \approx \frac{1}{4\pi} \int_{\alpha=0}^{2\pi} \int_{E=0}^{E_0} \frac{D^2 - D_P^2}{E^2} dE d\alpha, \tag{8.72}$$

where we have assumed that $R \sin\psi \approx 2R \sin\frac{\psi}{2} \approx R\psi \approx E$ and $R^2 d\sigma \approx EdEd\alpha$, and E is the surface distance between computation and integration points (truncated at distance E_0). Furthermore, by assuming that D varies by the polynomial

$$D = D_P + E\left(D_x \cos\alpha + D_y \sin\alpha\right) + \frac{E^2}{2}\left(D_{xx}\cos^2\alpha + D_{xy}\sin 2\alpha + D_{yy}\sin^2\alpha\right), \tag{8.73}$$

where all D_i and D_{ij} are horizontal derivatives in the x- and y-directions of D, it follows from Eq. (8.72) that:

$$I \approx \frac{E_0}{8}\left[D_P(D_{xx} + D_{yy}) + 2(D_x^2 + D_y^2)\right]. \tag{8.74}$$

The horizontal derivatives of D can be regarded as unknown coefficients in the surface polynomial representation of Eq. (8.73), and they can thus be determined by least squares from the (approximately) known values of D of the grid around the computation point.

It should be noted that the second-order solution above can also be determined more directly from Eq. (8.73) when replacing δD by D and expanding the Taylor series to second-order.

8.5.4 Additive Gravity Corrections

To solve the gravimetric problems of isostasy for finding the Moho parameters, the gravitational contributions of all known mass-density contrasts within the Earth's crust should be modelled and subsequently removed from observed gravity data. Moreover, inhomogeneous density structures within the mantle lithosphere and sub-lithosphere mantle (including the core-mantle geometry) will be taken into consideration, provided that reliable data of the global mantle-density structure are available. The global correlation of various gravity-field quantities with the Moho geometry can be investigated to find proper corrections to the gravity data. A maximum correlation is attained when using the gravity disturbances corrected for the gravitational contributions of topography and mass-density contrasts of ocean, ice and sediments (Tenzer et al. 2009; Bagherbandi et al. 2013).

The refined Bouguer gravity disturbances δg^R are firstly calculated from the surface gravity disturbances δg according to the following scheme (Tenzer et al. 2012; Bagherbandi et al. 2013):

$$\delta g^R = \delta g - \left(\delta g^{\mathrm{t}} + \delta g^{\mathrm{b}} + \delta g^{\mathrm{i}} + \delta g^{s}\right), \tag{8.75}$$

where δg^t is the topographic gravity correction, and δg^b, δg^i and δg^s are, respectively, the stripping gravity corrections for the ocean (bathymetry), ice and sediment.

By applying uniform mathematical formalism the gravity corrections of density contrast within the Earth's crust, δg^t, δg^b, δg^i and δg^s can be computed. It utilizes the expression for the gravitational attraction g (defined as a negative radial derivative of the respective potential V; i.e. $g = -\partial V/\partial r$) generated by an arbitrary volumetric mass layer with a variable depth and thickness while having laterally distributed vertical mass-density variations. The topographic, bathymetric, ice, sediment and consolidated crust coefficients can be generated with a spherical resolution complete to degree n of spherical harmonics and up to the third-order terms of a binomial expansion according to Tenzer et al. (2012, 2014). Simply, the

spherical harmonic analysis for determination of the effects of major known crustal-density structures are given below:

$$c^q_{\mathrm{nm}} = \frac{3}{(2n+1)\rho_e}\left[\frac{\left(\rho^q L_q\right)_{nm}}{R} + \frac{n+2}{2}\frac{\left(\rho^q L_q^2\right)_{nm}}{R^2} + O_{nm}\right], \tag{8.76a}$$

and

$$\left(\rho^q L_q^j\right)_{nm} = \frac{1}{4\pi}\iint_{\sigma} (density)^q \left(thickness^j\right)^q Y_{nm}(Q)\, d\sigma, \quad j = 1 \text{ and } 2 \tag{8.76b}$$

where ρ^q and L_q are density and thickness of the layers. Here q denotes either the topography, bathymetry, ice or sediment. O_{nm} is the effect of higher-order terms that can be neglected. Finally, the gravity disturbance corrections are given by:

$$\delta g^q = \frac{GM}{R^2}\sum_{n=0}^{n_{\max}}(n+1)\sum_{m=0}^{n} c^q_{nm} Y_{nm}(P). \tag{8.77}$$

The refined isostatic gravity disturbance δg^{i} is defined as (Tenzer and Bagherbandi 2012; Sjöberg 2013)

$$\delta g^{\mathrm{i}}(r,\Omega) = \delta g^{\mathrm{R}}(r,\Omega) + A_C(r,\Omega), \tag{8.78}$$

where δg^{R} is the refined Bouguer gravity disturbance (i.e. the topography-corrected and bathymetry-stripped gravity disturbance), and A_C is the isostatic compensation attraction.

The global maps of the topographic and crust components (stripping) gravity corrections and step-wise-corrected gravity disturbances are shown in Fig. 8.20. The statistics of the corrected gravity disturbances are summarized in Table 8.3. The GOCO-03S coefficients (Mayer-Guerr et al. 2012) and the parameters of the GRS-80 normal gravity field (Moritz 2000) were used to compute the gravity disturbance.

8.5.4.1 Maximum Gravity Signature of the Moho Geometry

In this part, a correlation spectrum analysis of the step-wise corrected (stripped) gravity disturbance with crustal thickness is presented. The idea is to compile the gravity-field quantities generated by the Earth's crustal structures and to investigate their spatial and spectral characteristics and their correlation with the crustal geometry in the context of the gravimetric Moho determination. This analysis can help to find the upper-most degree/resolution of Moho using the

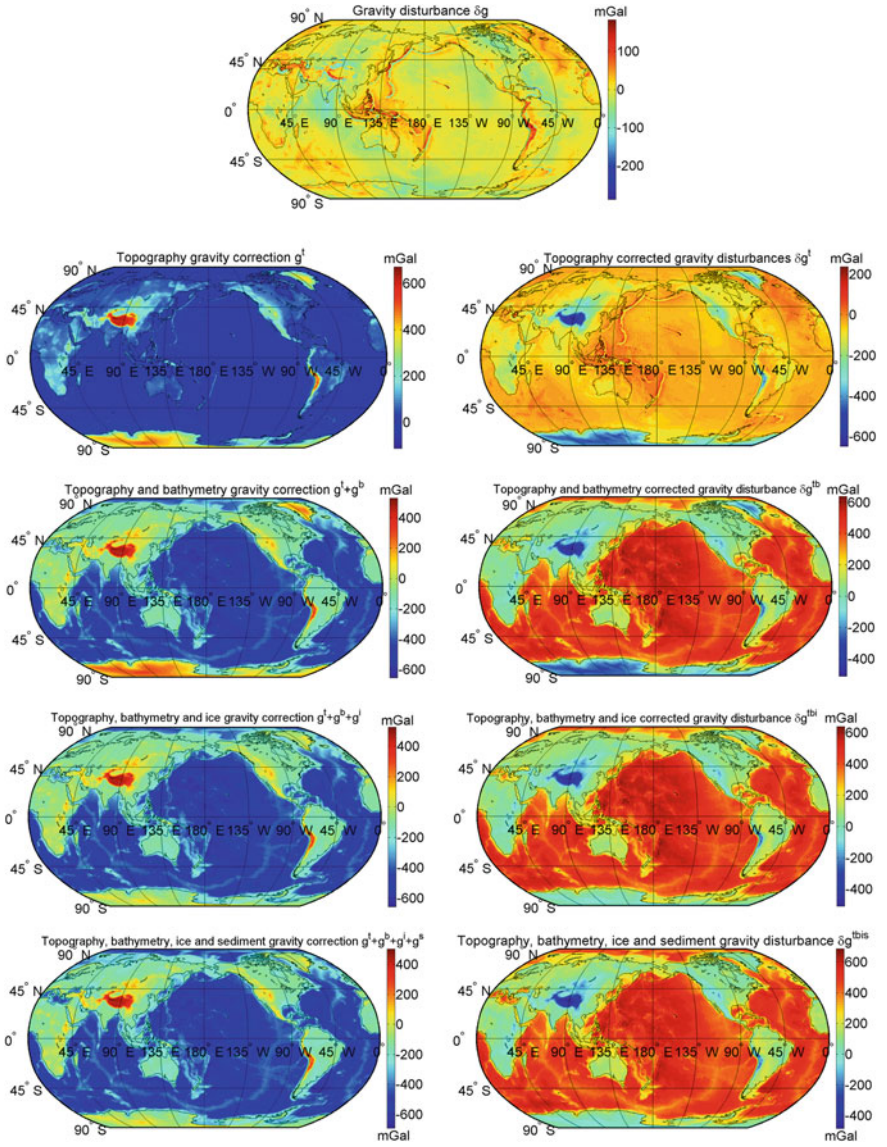

Fig. 8.20 Global maps of the gravity corrections and the step-wise corrected gravity disturbances computed globally on a 1 × 1 arc-deg grid of surface points with a spectral resolution complete to degree and order 180 of spherical harmonics. Unit: mGal

isostatic-gravimetric model. For this purpose, the degree variances and cumulative degree variances of the (step-wise) corrected gravity disturbances are compared.

The degree-correlation coefficients are calculated to assess a spectral harmonic correlation, for example, between the corrected gravity disturbances and the crustal

Table 8.3 Statistics of the step-wise corrected gravity disturbances computed globally on a 1 × 1 arc-deg grid of surface points with a spectral resolution complete to degree and order 180 of spherical harmonics: the GOCE03S gravity disturbances δg; the topography-corrected gravity disturbances δg^{t} ($=\delta g - \delta g^{t}$); the topography-corrected and bathymetry-stripped gravity disturbances δg^{tb} ($=\delta g - \delta g^{t} - \delta g^{b}$); the topography-corrected and bathymetry- and ice-stripped gravity disturbances δg^{tbi} ($=\delta g - \delta g^{t} - \delta g^{b} - \delta g^{i}$); the topography-corrected and bathymetry- and ice- and sediment-stripped gravity disturbances δg^{tbis}($=\delta g - \delta g^{t} - \delta g^{b} - \delta g^{i} - \delta g^{s}$) (including intermediate values of $\delta g^{tbi\ uS}$—after applying the upper-sediment stripping correction to δg^{tbi} and$\delta g^{tbi\ uSmS}$—after applying the upper- and middle-sediment stripping corrections to δg^{tbi})

Gravity disturbance	Min [mGal]	Max [mGal]	Mean [mGal]	STD [mGal]
δg	−229	257	−1	30
δg^{t}	−648	167	−72	107
δg^{tb}	−511	634	260	233
δg^{tbi}	−508	638	285	202
$\delta g^{tbi\ uS}$	−494	664	319	203
$\delta g^{tbi\ uSmS}$	−487	669	330	200
δg^{tbis}	−486	669	331	200

geometry. For a correlation between the gravity disturbances δg and the Moho depths D, we have:

$$r_n = \frac{\sum_m (T_{nm})(D_{nm})}{\sqrt{\sum_m (T_{nm})^2}\sqrt{\sum_m (D_{nm})^2}}, \tag{8.79}$$

where T_{nm} and D_{nm} are the harmonic coefficients of the disturbing disturbances and Moho depths.

The spectral correlation of the (step-wise) corrected gravity disturbance with the Moho geometry (up to degree and order 180) is shown in Fig. 8.21. CRUST1.0 is used for the correlation analysis. The figure shows that all investigated types of the gravity disturbances comprise the largest gravity correlation at long wavelengths, while, at higher-frequencies, the correlations almost monotonously attenuate. Except for the ice-gravity correction, the application of the gravity corrections increases the correlation at almost the entire investigated gravity spectrum (up to degree and order 180). This spectral behaviour corresponds to the spatial characteristics of the corrected-gravity disturbances, which ranges of values mostly increase after a subsequent application of each individual gravity correction. The use of the ice-gravity correction, on the other hand, substantially reduces (in the absolute sense) large negative values of the gravity disturbances in polar areas of Greenland and Antarctica. The largest changes in degree variances were caused by applying the topographic and bathymetric gravity corrections. The application of the topographic gravity corrections changed the gravity spectrum especially at the long-to-medium wavelengths up to degree of 60, while changes due to applying the bathymetric gravity correction are seen up to degree 100 (Tenzer et al. 2014). The maximum correlation is at the long wavelengths (more than 0.9 below degree 20), while the

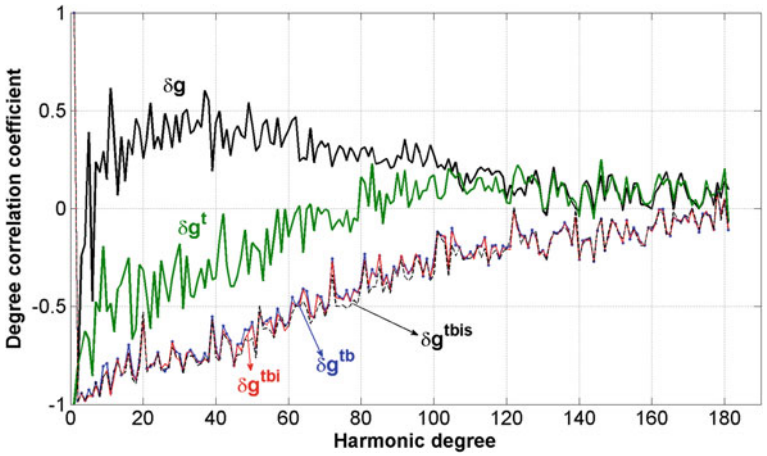

Fig. 8.21 Spectral correlation of the step-wise corrected gravity disturbances with the Moho geometry (complete to degree and order 180 of spherical harmonics)

Table 8.4 The spatial correlations of the step-wise corrected gravity disturbances with the solid Earth topography and Moho depth (CRUST1.0)

Gravity disturbances	Correlation with	
	Solid topography	Moho depth
δg	−0.01	−0.09
$\delta g - g^{t}$	−0.55	−0.68
$\delta g - g^{t} - g^{b}$	−0.89	−0.95
$\delta g - g^{t} - g^{b} - g^{i}$	−0.92	−0.97
$\delta g - g^{t} - g^{b} - g^{i} - g^{s}$	−0.94	−0.97

correlation almost monotonously attenuates with an increasing frequency to less than 0.25 (at degree 180).

The spatial correlation of the step-wise corrected gravity disturbances with the solid-Earth topography and the Moho geometry is given in Table 8.4. GOCO-03S gravity disturbances are not significantly correlated with the solid topography (−0.01), and the correlation with the Moho depth (CRUST1.0) is only −0.09. Remarkably, the Earth's gravity field has a slightly higher (absolute) correlation with the Moho geometry than with the solid topography (Tenzer et al. 2014). This indicates that, at the investigated gravity spectrum (up to degree 180 of spherical harmonics), the gravity signatures of crustal- and mantle-density heterogeneities are more pronounced in the Moho geometry. The higher-frequency spectrum of the Earth's gravity field is, on the other hand, dominated by the pattern of the terrain and ocean floor relief. The application of the topographic and bathymetric gravity corrections significantly increased the correlation of the gravity disturbances with the solid topography (−0.89), as well as with the Moho geometry (−0.95). Application of the additional ice-, sediment- and consolidated-crust gravity

corrections further increased the correlation of the gravity disturbances to −0.95 (with the solid topography) and −0.98 (with the Moho geometry). These gravity disturbances thus comprise a maximum gravity signature of the Moho geometry and show that, up to degree 180, it is enough to determine the Moho depth using gravity data. However, these gravity disturbances contain also the gravity signal of unmodelled mantle heterogeneities (including the core-mantle boundary zone; Peltier (2007); Bagherbandi and Sjöberg (2013)).

8.5.5 *The Non-isostatic Effects*

The crustal thickness can be determined based on gravimetric-isostatic and seismic models. Modelling crustal thickness by the former model suffers from some problems. The isostatic assumption for compensating the topographic potential is incomplete, as there are other effects which should be considered. It is almost impossible to distinguish the crustal- and mantle-gravity field and geothermal modelling without additional data on the crustal structure (e.g. Kaban et al. 2004; Artiemeva 2006; Artiemeva et al. 2006; Tesauro et al. 2008). In reality, Moho is not only formed by isostasy, but there are other causes, which affect the crustal-thickness estimation. Using the isostatic hypothesis for determining the depth of crust causes some disturbing signals, non-isostatic effects (NIEs), which influence the crustal thickness determination. There are two types of NIEs:

1. Mass density variations below sea level that are not related with isostasy. This category includes part of the Moho geometry formed by other processes than isostasy.
2. Remaining (unknown) mass-density variations in the topography, which were not considered in forming the refined Bouguer disturbances.

Hence the NIEs are the gravitational effects that are caused by the deviation of Moho geometry from its isostatic model. Major parts of the long-wavelengths of the geopotential are caused by density variations in the Earth's deep mantle and core/mantle topography variations (Martinec 1994). According to Bagherbandi and Sjöberg (2013), the isostatic assumption for compensating the topographic potential is inadequate, and other effects should be considered. Therefore, the assumptions of Moritz (1990) and Sjöberg (2009) for Vening Meinesz' inverse problem should be corrected by this effect. Bagherbandi and Sjöberg (2013) presented the following method:

1. Determine the potential harmonic coefficient of the NIE by:

$$c_{nm}^{NIE} = c_{nm}^{\text{Seismic}} - c_{nm}^{VMM}, \tag{8.80}$$

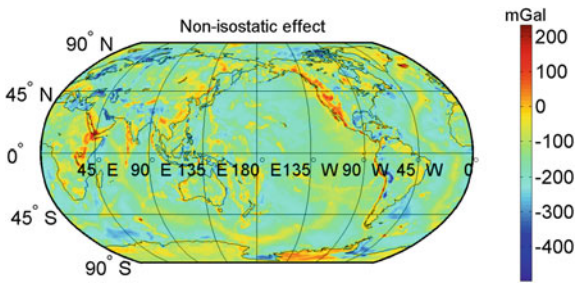

Fig. 8.22 Non-isostatic effects. Unit: mGal

where spherical harmonic coefficients c_{nm}^{j} (j = VMM, Seismic) are given by:

$$c_{nm}^{j} \approx \frac{3}{(2n+1)\rho_e}\left[\frac{\left(\Delta\rho\left(D_j - D_0\right)\right)_{nm}}{R} + \frac{(n+2)\left(\Delta\rho(D_0^2 - D_j^2)\right)_{nm}}{2R^2} + \frac{(n+2)(n+1)}{6}\frac{\left(\Delta\rho(D^3 - D_0^3)\right)_{nm}}{R^3}\right], \tag{8.81}$$

and $\rho_e \approx 5.5$ g/cm^3 is the mean density of the Earth's mass. The spherical harmonic coefficients of $\Delta\rho\left(D_j - D_0\right)$, $\Delta\rho(D_0^2 - D_j^2)$ and $\Delta\rho(D_j^3 - D_0^3)$ are shown by $\left(\Delta\rho\left(D_j - D_0\right)\right)_{nm}$, $\left(\Delta\rho(D_0^2 - D_j^2)\right)_{nm}$ and $\left(\Delta\rho(D_j^3 - D_0^3)\right)_{nm}$.

2. The NIE correction to the Bougure gravity disturbance becomes:

$$\delta g^{NIE} = \frac{GM}{R^2}\sum_{n=0}^{n_{max}}(n+1)\left(\frac{R}{r_p}\right)^{n+2}\sum_{m=-n}^{n} c_{nm}^{NIE} Y_{nm}(P). \tag{8.82}$$

Figure 8.22 shows the NIE globally due to the long-wavelength features of the gravity field. The maximum deviations from isostasy occur at currently active plate boundaries. Positive NIEs in the upper mantle under the continents are found in Iceland and East Africa, and in the vicinity of the axes of the mid-ocean ridges, indicating a compensation mechanism due to thermal and compositional density (Kaban et al. 2004). The most pronounced negative NIEs in the upper mantle are found in areas with post-glacial rebound (Fennoscandia and Eastern Canada) and along the south-west coast line of South America and in West Antarctica. In the presence of the NIEs, the obtained isostatic gravity anomaly, e.g. from Eq. (8.64a), is not suitable for determining crustal thickness, and it needs a correction. The pure isostatic-gravity anomaly should be an isostatic- gravity anomaly due to the attraction of the crust and its isostatic compensation only, which is our desire in the ideal VMM model. Comparing the results presented by Kaban et al. (2010) and the NIEs (in this study) shows that both techniques for determining mantle gravity field are consistent.

8.5.6 Thermal-Pressure Effect Due to Lithosphere-Mantle Density

Several methods have been proposed for oceanic lithospheric mantle-temperature calculation and consequently also for the lithosphere thermal-gravity anomaly, for example, by McKenzie (1978), Parsons and Richter (1980), Bouhifd et al. (1996), McKenzie et al. (2005), Afonso et al. (2008) and Bai et al. (2014). Thermal isostasy is the geodynamic process whereby regional variations in the lithospheric thermal regime cause changes in crustal thickness. The changes result from variations in rock density in response to thermal expansion. Therefore, the lithospheric mantle density is not constant, and the density variation should be taken into account for gravity-inversion studies.

The density of the lithospheric mantle in the thermal regime would be reduced by the effect of thermal expansion. On the other hand, the density could also be increased when affected by the pressure-driven compression due to the loading materials. If the buried depth of the lithospheric mantle unit is z, then its density could be evaluated by:

$$\rho_z = \rho_0[-\alpha(T_z - T_0) + \beta_{T_z}(P_z - P_0)], \tag{8.83}$$

where ρ_0 (typically set to 3.3 g/cm^3) is the lithospheric mantle density at normal temperature T_0 (typically set to 273 K) and normal pressure P_0 (typically set to 101 kPa, the standard atmospheric pressure), α is the thermal expansion coefficient (we set its value constantly to $3.28 \times 10^{-5} K^{-1}$ according to former studies by Bai et al. (2014) and Chappell and Kusznir (2008)), β_{T_z} is the pressure-driven compressibility coefficient relevant to the temperature, T_z and P_z are the temperature and pressure at depth z, respectively.

8.5.6.1 Temperature Field

According to the pure shear model (McKenzie 1978), the temperature at depth z could be estimated by:

$$T_z = T_1\left(1 - \frac{(a - \mathrm{z})}{a} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\left[\frac{\gamma}{n\pi}\sin\frac{n\pi}{\gamma}\right]\exp\left[\frac{-n^2 t}{\tau}\right]\sin\frac{n\pi(a - \mathrm{z})}{a}\right), \tag{8.84}$$

where T_1 (typically set to 1060 K) is the boundary temperature of lithosphere and asthenosphere, a (typically set to 125 km; cf. Chappell and Kusznir 2008 and Bai et al. 2014) is the equilibrium lithosphere (plate) thickness, τ (typically set to 62.8 Myr;

Bai et al. 2014) is the lithosphere cooling thermal-decay constant, t is crustal age, γ is lithosphere stretching factor (McKenzie et al. 2005). The oceanic crustal age could be estimated based on magnetic lineation and the continental crustal age is set constantly to 300 Ma in this study (see Sect. 8.5.6.4). The lithospheric-stretching factor could be approximated by the crustal-stretching factor, which is infinite for the oceanic lithosphere (Chappell and Kusznir 2008). The initial values of continental crustal-stretching factors, which are needed for thermal modelling, are based on crustal thickness mapped by gravity inversion. The initial crustal thickness is considered without thermal-expansion and pressure-compression effects, and these effects can be calculated iteratively.

8.5.6.2 Pressure-Driven Compressibility Coefficient

The bulk modulus (K_T) could be simulated by a function of temperature, T, given by (Kroll et al. 2012) as:

$$K_T = 127.97 - 0.0232 \times (T - 300), \tag{8.85}$$

and its inverse is the pressure-driven compressibility coefficient (β_{T_z}):

$$\beta_{T_z} = 1/K_T. \tag{8.86}$$

8.5.6.3 Pressure Field

The hydrostatic pressure at depth z could be calculated based on the definition of lithospheric pressure P_z as (Afonso et al. 2008):

$$P_z = \int_0^z \rho(\mathrm{z}')\mathrm{dz}'. \tag{8.87}$$

In the space domain, the gravity disturbance at point P on the sphere of the Earth due to thermal and pressure effects could be calculated by:

$$\delta g^{TE}(P) = G \iint_{\sigma} \int_{R-z}^{R} \frac{(\rho_z - \rho_0)r}{l_P^3} dr d\sigma. \tag{8.88}$$

where ρ_z is given by Eq. (8.83). This implies that both corrections δg^{NIE} and δg^{TE} should be added as corrections to the isostatic gravity- disturbance (δg^{i}), yielding

the (theoretically) pure isostatic gravity disturbance equation for Moho determination (see the application of δg^{NIE} and δg^{TE} in Fig. 8.23):

$$\delta g^{I}(P) = \delta g^{R}(P) + \delta g^{NIE}(P) + \delta g^{TE}(P) + A_C(P) = 0. \tag{8.89}$$

8.5.6.4 Global and Regional Crustal Thicknesses

Figure 8.23 shows the global crustal thicknesses estimated from the VMM model after correction for the NIE and thermal effects. In most areas, one observes a significant improvement by adding the corrections. For example, considerable improvements can be seen in the areas with post-glacial rebound (such as Fennoscandia and Hudson Bay in Canada). In areas with existing huge ice masses, e.g. Antarctica and Greenland, the estimates of crustal thickness are also improved. The details in Fennoscandia and South America can be seen in the sub-figures. These areas are selected because (a) in Fennoscandia, the seismic Moho strongly

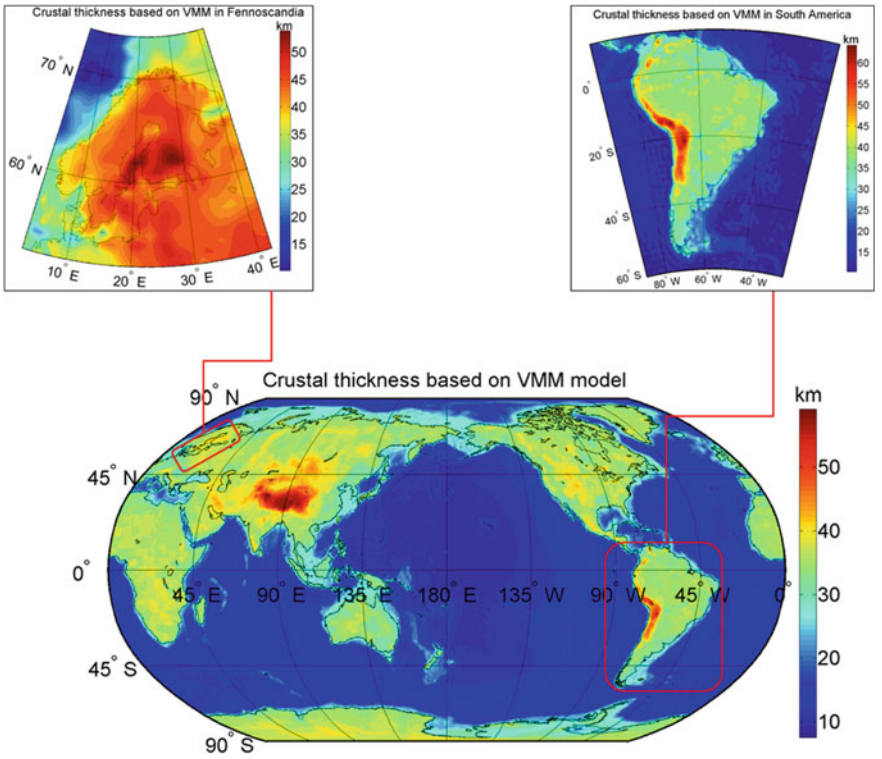

Fig. 8.23 Crustal thickness based on VMM model after stripping gravity corrections due to the topographic, ocean (bathymetry), ice and sediments and non-isostatic and thermal effects with a resolution of 1 × 1 arc-degree. Unit: km

deviates from the gravimetric one, and the area contains a significant non-isostatic effect by GIA, which complicates the estimation of the crustal thickness by the isostatic modelling technique, and (b) in South America, there are huge topographic masses in Peru and Chile that abruptly change to a deep ocean trench to the west. For this purpose, the following parameters and models are considered. The thermal-pressure correction is calculated using the oceanic crustal-age grid (Müller et al. 2008; Seton et al. 2012), solid-earth topographic heights from ETOPO1 data, total sediment thickness of the world's oceans and marginal seas by Divins (2003) and the global sediment-thickness grid onshore by Dziewonski and Anderson (1981); the bathymetry- and sediment-thickness grids are used for Moho surface estimation. The parameter z is assumed to be the depth between Moho to 125 km (bottom of the lithosphere) in Eqs. (8.83) and (8.84). This also means that the temperature and density are determined for a part of the mantle (deeper than Moho, but shallower than 125 km). Here Moho depth should be calculated iteratively. The initial value for Moho is based on gravity inversion without considering thermal and pressure effects. Therefore, based on this value, the mantle temperature and density could be modelled, and then a new Moho depth can be obtained by taking thermal and pressure effects into consideration. This process should be performed interactively to determine the final Moho geometry.

8.5.6.5 Crustal Thickness in Fennoscandia

The Moho map in Fennoscandia shows that the deepest Moho boundaries are in Finland and in the Baltic Sea, which can be revealed in a geological map. The maximum Moho depressions beneath the Gulf of Bothnia (55.3 km) and Southern Finland (55.1 km) are clearly observed in the gravimetric-isostatic model and reach to the depths of 4.8 km deeper and 3 km more shallow than seen in the CRUST1.0. The eastward deepening from the Norwegian coast to Sweden and then into the Baltic Sea and Finland is also observed.

According to the geological map in Fennoscandia, shown in Fig. 8.24a, various types of geological provinces (units) can be observed. A zone of high-density values can be seen along the boundary between Sweden and Norway, where the bulk part of the Scandinavian Mountain Chain (the Scandes) is located. This zone is believed to be a lithospheric transition zone discussed in previous studies (e.g. Medhus et al. 2012; Gradmann et al. 2013). The zone is located between the lower and upper mantle. A major unit located both within and to the east of Fennoscandia is the Trans-Scandinavian Igneous Belt (TIB), which is an outstanding feature that has an important influence on the gravity field due to its generally low density in the shallow crust (Henkel and Eriksson 1987; Maupin et al. 2013). By dividing the mass of each block into the crustal root times $\Delta\rho$ ($\rho H/(\Delta\rho t_l)$), the compensation-ratio will be achieved. The ratio is probably introduced by the transition zone in Fennoscandia. We believe that the compensation-ratio reflects the density variation in the upper mantle, and the masses above sea level can probably be a good indicator of such transition zones and should be tested in some other study areas.

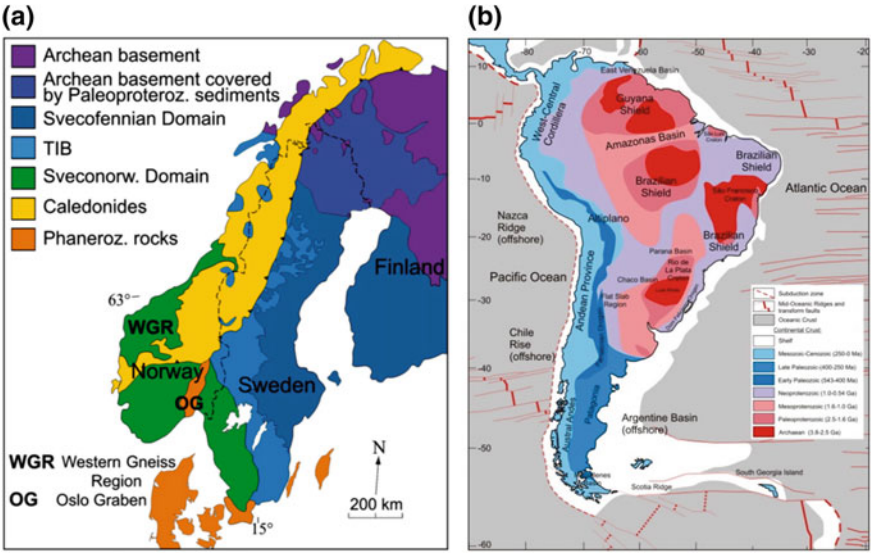

Fig. 8.24 **a** Simplified geological map of Fennoscandia after Gorbatchev (2004) and Maupin et al. (2013). TIB is a commonly used acronym for "Transscandinavian Igneous Belt", and **b** main geological provinces of South America (after Chulick et al. 2013; Gurbanov and Mooney 2012)

A correlation between the compensation ratio and geological province can be seen at the boundary between Norway and Sweden (especially in the Caledonides zone). The compensation-ratios denote the strengths of the various crustal units as the lower ratio corresponding to the stronger geological unit. According to Prasanna et al. (2014), the strength of every geological unit is a function of its lithology (that is, density) and thermal state, which correspond to the lithological properties of the various crustal plates. The compensation-ratio is close to zero in the oldest parts of the Baltic Shield, where the lowest heat-flow rates are observed (Bungum et al. 1980).

An alternative comparison can be performed between the VMM model and the point-wise seismic Moho depth data compiled by Luosto (1991) in Fennoscandia (see Fig. 8.25). Luosto's data was compiled in 308 points in the Fennoscandian shield. A comprehensive review of the seismic data collected in Fennoscandia has been presented by Luosto (1997). He presented the main models of seismic profiles covering Fennoscandia from the south to the north (Eken et al. 2008). The results of those studies indicate that the Moho deepens towards the centre of Fennoscandia, although there is no apparent corresponding topographical signature. Other interpretations of the Moho depth in Fennoscandia, such as tectonic implications, were discussed by Korja (1995), Korja et al. (1993) and Luosto (1997). The differences between the results of Luosto (1997) and VMM vary from −12.5 to 10.4 km with a mean value of 0.5 km and RMS of 4.7 km. The differences are less than 3 km in 67% of points (also 83% less than 5 km). The thickest crust can be observed in the

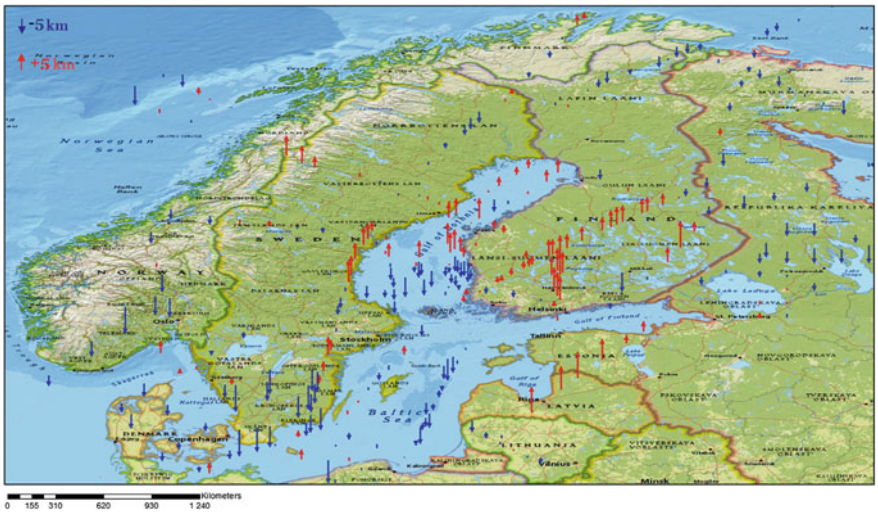

Fig. 8.25 Difference between crustal thicknesses obtained from seismic data compiled by Luosto (1997) and the VMM model after stripping gravity corrections. *Red/blue arrows* show positive/negative differences (see the scale in the *upper left corner*). Unit: km

Gulf of Bothnia and south of Finland, where the maximum differences between the two models occurs along a line between Söderhamn (in Sweden), Pori, Tampare and Mikkeli (in Finland). Northern Fennoscandia shows a rather uniform crustal thickness of around 40–45 km, which agrees well with the average profiling results in Fennoscandia. These results are comparable with the results presented by Bungum et al. (1980). There is complete compensation in the mountainous area located in the south-western part of Norway (based on both Airy and Pratt models). A similar situation can be observed in the areas with no topography (or less topography), where deep Moho exist, e.g. in the Gulf of Bothnia and south of Finland (meaning that variable-density contrast plays more of a role).

8.5.6.6 Crustal Thickness in South America

Figure 8.23 (top-right) depicts the crustal thickness of South America estimated by the VMM method. The figure shows thicker crust (more than 50 km) along the Andes and thinner crust west of the Andes' thick craton. The average crustal thickness of the South American continent is 43.5 km, which is comparable with the 45.7 km presented by Chulick et al. (2013). The corresponding Moho depths of the CRUST1.0 for this area varies between 9 and 69 km, with an average and standard deviation of 22 and 13 km, respectively. The RMS of these differences between the VMM and CRUST1.0 Moho depths and point-wise seismic refraction data (in the Andes) are of the order of 4.7/4.5 km and of 2.8/4.2 km before and after

applying the thermal correction, respectively. The result shows that the VMM model works well and can be complementary data for seismic models.

Figure 8.24b depicts the main geological provinces of South America, illustrating the large sediment basins in the Argentine Basin and East Venezuela Basin (Shelf). The minimum sediment layers are located in old cratons such as the Guyana, Brazilian Shields and Luis Alves cratons (2.5–3.8 Ga age) and belong to Archaean time. They are the old and stable part of the continental lithosphere that are generally found in the interiors of tectonic plates, characteristically composed of ancient crystalline basement rock, which may be covered by younger sedimentary rock. Also, these regions are different from the cratons, being more geologically active and unstable. Two other significant sediment basins are observed: Famatinian Orogen and Amazonas Basin. Generally, Precambrian South America is predominantly Proterozoic in age (Cordani and Sato 1999), and has suffered from several phases of continental collision and subsequent breakup (Lloyd et al. 2010). The contribution of the NIEs (the effect of mantle and lower-mass layers) on the gravity data are large, varying between −488 and 290 mGal and with a standard deviation of 107 mGal. Understanding the non-isostatic processes defining the mantle activity and the current state of cratons are key issues for unravelling the Earth's interior. The gravity field of the mantle and lower-located masses offers a starting point for numerical modelling of deep-Earth structures and the main tool to investigate the structure of the mantle (Tesauro et al. 2008). In order to investigate the reasons for significant NIE values, Fig. 8.24 can be considered as a reference model for comparison with the estimated NIEs in Fig. 8.26.

The comparison shows that the significant NIE values are observed in mid-ocean ridges and transform faults in the north-east and south-west of the study region (in the Atlantic and Pacific Oceans). In the mid-oceans, the magma is rising from a chamber below, forming new ocean ridges, which spread away from the ridge. Therefore, the density of the crust is higher than the other parts, indicating that the gravity field of the mantle dominates. The second largest NIE in South America can be found in Andean Province, where there is Paleozoic structures of the age of

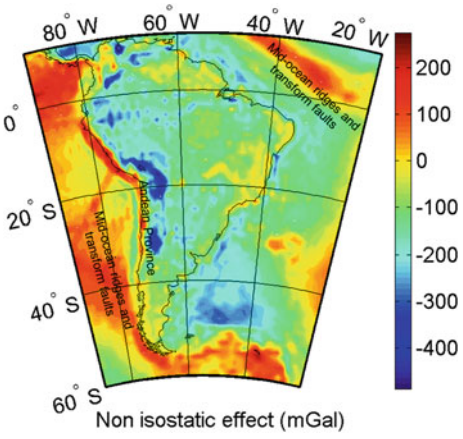

Fig. 8.26 Non-isostatic effects after removing thermal compensation from gravity disturbance. Unit: mGal

250–400 Ma. Franz et al. (2006) studied the evolution of the continental crust at the central Andes. Their studies show that the growth of the continental crust is closely linked to the phenomenon of subduction (still an ongoing process), and active continental margins (Franz et al. 2006). Therefore, the NIEs can also be used to study this phenomenon in geodynamics and in the behaviour of entire geological systems. Generally, the positive NIE values are in the oceans and the negative ones in continental areas. Hence, without considering the NIEs, there is over-compensation in the Andes region, but these effects help in reaching the isostatic balance.

8.5.7 Combined Moho Determination

In most of isostatic hypotheses, one of the Moho parameters, Moho depth or Moho density contrast (MDC), is constant. However, assuming constant MDC or Moho depth is not correct. One of the solutions that can be used in both gravimetric-isostatic and seismic crustal-thickness models is a combined least-squares adjustment of the two parameters (Sjöberg and Bagherbandi 2011). If we assume that the density contrast is variable, Eq. (8.65a) becomes:

$$b(P) = R \iint_{\sigma} \Delta\rho K(\psi, s) d\sigma, \tag{8.90}$$

where:

$$b(P) = -\left[\delta g^{R} + A_{C0}\right]/G. \tag{8.91}$$

The left hand-side can be expanded as a Taylor series (limited to second order of $\tau = D/R$; Sjöberg 2009):

$$\begin{aligned} R \iint_{\sigma} \Delta\rho K(\psi, s) d\sigma &= R \sum_{n=0}^{\infty} (n+1) \iint_{\sigma} \Delta\rho (\tau - \frac{n+2}{2}\tau^2) P_n(\cos\psi) d\sigma \\ &= 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{n+1}{2n+1} \left[(\Delta\rho D)_{nm} - \frac{n+2}{2R} \left(\Delta\rho D^2\right)_{nm} \right] Y_{nm}(P), \end{aligned} \tag{8.92}$$

and, in view of Eq. (8.90) and the spherical harmonic series of b, one thus obtains the spectral equation:

$$\frac{4\pi(n+1)}{2n+1}[\,] = b_{nm}, \quad \text{or} \quad [\,] = \frac{2n+1}{4\pi(n+1)} b_{nm}, \tag{8.93}$$

where the bracket is the same as in Eq. (8.92). By summing up the spectral solutions $\Delta\rho D$, we thus obtain from the last term of Eq. (8.93):

$$(\Delta\rho D)_P = \sum_{n=0}^{\infty}\sum_{m=-n}^{n}\left[\frac{2n+1}{4\pi(n+1)}b_{nm} + \frac{n+2}{2}\frac{(\Delta\rho D^2)_{nm}}{R}\right]Y_{nm}(P), \tag{8.94}$$

which can also be written in a closed form as (Sjöberg 2009):

$$\begin{aligned}\Delta\rho_P \approx \Bigg[&\frac{b_P}{2\pi D_P} - \frac{1}{(4\pi)^2 D_P}\iint_{\sigma} H(\psi)b(Q)d\sigma + \frac{(D\Delta\rho)_P}{R} \\ &- \frac{1}{32\pi R T_P}\iint_{\sigma}\frac{(\Delta\rho D^2)_Q - (\Delta\rho D^2)_P}{\sin^3(\psi/2)}d\sigma\Bigg]\end{aligned} \tag{8.95}$$

where

$$H(\psi) = \sum_{n=0}^{\infty}\frac{2n+1}{n+1}P_n(\cos\psi) = \operatorname{cosec}(\psi/2) + \ln[1+\operatorname{cosec}(\psi/2)]. \tag{8.96}$$

Here, the first two terms of Eq. (8.95) stem from the first term on the right hand-side of Eq. (8.94). With reference to Heiskanen and Moritz (1967, Sect. 1–18), the last term of Eq. (8.94) yields the integral plus the term ahead of it in Eq. (8.95). Equation (8.95) lends itself to iteration of the density contrast for known Moho depths. Unfortunately, the last integral of the equation has a singularity. To eliminate its influence, the near-zone integration area can be approximated by a plane. The last integral of Eq. (8.95) should only be significant in the near-zone of the computation point. Hence, by introducing the notations $u = D^2\Delta\rho$ and $u_P = (D^2\Delta\rho)_P$ and following the derivations of Heiskanen and Moritz (1967, pp. 121–122), it can be approximated by the plane integral, truncated to the polar distance E_0 from the computation point P similar to Eq. (8.72):

$$I \approx \frac{1}{4\pi}\int_{\alpha=0}^{2\pi}\int_{E=0}^{E_0}\frac{u-u_P}{E^3}\,EdEd\alpha, \tag{8.97a}$$

where (E,α) are the polar distance (truncated at distance E_0) and azimuth. Representing u by the truncated Taylor series

$$u \approx u_P + Eu_x\cos\alpha + Eu_y\sin\alpha + \frac{E^2}{2}\left(u_{xx}\cos^2\alpha + 2u_{xy}\sin\alpha\,\cos\alpha + u_{yy}\sin^2\alpha\right), \tag{8.97b}$$

Equation (8.97a) becomes:

$$I \approx \frac{E_0}{8}\left(u_{xx} + u_{yy}\right). \tag{8.97c}$$

The horizontal derivatives of u in Eq. (8.97c) can be determined in a least-squares adjustment from preliminary observations of u in a grid around P.

8.5.7.1 Approximate Solution by Spherical Harmonics

If the left hand-side of Eq. (8.94) is also expanded as a series in spherical harmonics, we arrive at the spectral equation:

$$(\Delta\rho D)_{nm} = \frac{2n+1}{(n+1)4\pi} b_{nm} + \frac{n+2}{2R}\left(\Delta\rho D^2\right)_{nm}. \tag{8.98}$$

As the last term of this equation is much smaller than the remaining terms, we may approximate $(\Delta\rho D^2)_{nm}$ by $D_0(\Delta\rho D)_{nm}$, and we obtain from Eq. (8.98):

$$(\Delta\rho D)_{nm}\left(1 - \frac{n+2}{2}\frac{D_0}{R}\right) \approx \frac{2n+1}{4\pi(n+1)} b_{nm}, \tag{8.99}$$

with the solution for $\Delta\rho$:

$$\Delta\rho_P \approx \frac{1}{4\pi D}\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}\frac{2n+1}{n+1}\frac{1}{1-\frac{n+2}{2}\frac{D_0}{R}} b_{nm} Y_{nm}(P), \tag{8.100a}$$

or

$$\begin{aligned}\Delta\rho_P &\approx \frac{1}{4\pi D_P}\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}\frac{2n+1}{n+1}\left(1 + \frac{D_0}{R}\frac{1}{\frac{2}{n+2} - \frac{D_0}{R}}\right) b_{nm} Y_{nm}(P) \\ &= \frac{b_P}{2\pi D_P} - \frac{1}{4\pi D_P}\sum_{n=0}^{n_{\max}}\sum_{m=-n}^{n}\left(\frac{1}{n+1} - \frac{D_0/R}{2/(n+2) - D_0/R}\right) b_{nm} Y_{nm}(P).\end{aligned} \tag{8.100b}$$

Hence, if D_P is known, the MDC can be estimated from the spectrum of b_P.

8.5.7.2 Estimating Moho Depth and Moho Density Contrast Simultaneously

To solve for both density contrast $\Delta\rho$ and Moho depth D, with independent (a priori) estimates, d and κ of those parameters can be formulated as an adjustment by elements. Then the system of linearized observation equations becomes:

$$\mathbf{AX}=\mathbf{L}-\boldsymbol{\varepsilon}, \tag{8.101a}$$

where

$$\mathbf{A}=\begin{bmatrix}\kappa & t\\ 0 & 1\\ 1 & 0\\ 0 & 1\end{bmatrix},\quad \mathbf{X}=\begin{bmatrix}dD\\ d\kappa\end{bmatrix}\quad\text{and}\quad \mathbf{L}=\begin{bmatrix}l_1-d\kappa\\ l_2-\kappa\\ l_3-d\\ l_4-\kappa\end{bmatrix}. \tag{8.101b}$$

where l_1, l_2, l_3 and l_4 are the observations, provided as preliminary estimates of $D\Delta\rho$, $\Delta\rho$, D and $\Delta\rho$, respectively. l_1 and l_2 are obtained from gravity data and l_3 and l_4 from seismic data. Here dD and $d\kappa$ are the corrections to the *a priori* estimates of D and $\Delta\rho$. The least-squares solution of this system becomes:

$$\hat{\mathbf{X}}=\left(\mathbf{A^T Q^{-1} A}\right)^{-1}\mathbf{A^T Q^{-1} L}, \tag{8.102}$$

with the covariance matrix:

$$\mathbf{Q_{XX}}=\sigma_0^2\left(\mathbf{A^T Q^{-1} A}\right)^{-1}, \tag{8.103}$$

where matrix $\mathbf{Q}$ is the covariance matrix of the observations. Using the error propagation law to estimate the full covariance matrix becomes:

$$\mathbf{Q}=\begin{pmatrix}\sigma_1^2 & \sigma_1^2/d & 0 & 0\\ \sigma_1^2/d & \sigma_2^2 & 0 & 0\\ 0 & 0 & \sigma_3^2 & \sigma_{34}\\ 0 & 0 & \sigma_{34} & \sigma_4^2\end{pmatrix}, \tag{8.104}$$

where σ_1 and σ_3 are the standard errors of $D\Delta\rho$ and D, respectively, $\sigma_2^2=\sigma_1^2/d^2+\sigma_3^2(D\Delta\rho)^2\Big/d^4$, σ_4 is the standard error of $\Delta\rho$. σ_{34} is the covariance between l_3 and l_4 (as D and $\Delta\rho$ are obtained from the seismic data). The term $D\Delta\rho$ and its standard error (σ_1) are estimated by Eq. (8.94) (Sjöberg and Bagherbandi 2011) so:

$$\sigma_{D\Delta\rho}^2\approx\left(\frac{\gamma}{4\pi G}\right)^2\sum_{n,m}N_{nm}^2\sigma_{nm}^2+2\left(\frac{\gamma}{4\pi G}\right)^2\sum_{n,m}\sum_{k,l\neq n,m}N_{nm}N_{kl}\sigma_{nmkl}, \tag{8.105}$$

The variance of unit weight (σ_0^2) for two degrees of freedom can now be estimated by:

$$s_0^2 = \frac{\left(\mathbf{L}-\mathbf{A}\hat{\mathbf{X}}\right)^{\mathbf{T}}\mathbf{Q}^{-1}\left(\mathbf{L}-\mathbf{A}\hat{\mathbf{X}}\right)}{2}, \tag{8.106}$$

Finally the least-squares solution for the Moho depth and the MDC becomes:

$$\hat{D} = D + d\hat{D}, \tag{8.107}$$

and

$$\Delta\hat{\rho} = \Delta\rho + d\Delta\hat{\rho}, \tag{8.108}$$

where $d\hat{D}$ and $d\Delta\hat{\rho}$ are the estimated improvements in Eq. (8.102) to D and $\Delta\rho$, which are a priori values of the Moho depth and Moho density contrast.

The numerical application of the method that follows next is limited to only three observations, i.e. observation l_4 is not included. Here the approximate value for D (d) is set to 35 (13) km in continental (oceanic) areas, respectively. The approximate value (κ) of the MDC is set to 400 and 250 kg/m^3 in continental and oceanic regions, respectively. The observation vector $\mathbf{L}$ is composed of the observations $l_1 = D\Delta\rho$ as determined by Eq. (8.94), $l_2 = \Delta\rho$, given by Eq. (8.100a), and $l_3 = D$, given by the model CRUST1.0. The covariance matrix $\mathbf{Q}$ in Eq. (8.104) needs a full covariance matrix, as there is a significant correlation between l_1 and l_2. This is because l_1 and l_2 are estimated from the same data (namely the estimated Bouguer gravity disturbance). Another important assumption in our numerical study is related to the standard error of D. As the CRUST1.0 model is not provided with a standard error model, we follow Čadek and Martinec (1991) who assume 20% (6 km) and 10% (3 km) of the mean Moho depth as the standard error (σ_3) of the Moho depth in continental and oceanic regions, respectively. The result can be seen in Table 8.5.

Table 8.5 Statistics of global estimates of the MDC and Moho depth estimated by least-squares adjustment for 1° × 1° block data

Unit	Quantities	Max.	Mean	Min.	STD	RMS
kg/m^3	$\Delta\rho$	610	329.7	19.9	91.5	
	$d\Delta\hat{\rho}$	233.2	8.6	−311.6	58.6	
	$\Delta\hat{\rho}$	636.6	321.1	21	104.7	
	$C_{\Delta\hat{\rho}}$	270.8	44.7	0	34.3	
km	D	74.8	22.8	7.4	12.3	
	$d\hat{D}$	19.1	−0.9	−40.9	4.7	
	$\hat{D}$	85.9	23.8	6.2	14.1	
	$C_{\hat{D}}$	15.5	2.9	0.01	2.1	
	$\hat{D} - D^{CRUST1.0}$	40.9	0.9	−19.1	4.7	4.8

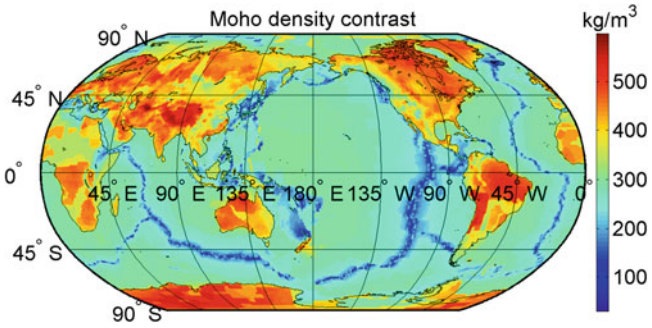

Fig. 8.27 Estimated MDC determined by combined model. Unit: kg/m^3

Figures 8.27 and 8.28 map the estimated MDC and combined Moho depth. Figure 8.27 shows that the maximum value (636 kg/m^3) is in Tibet, and the minimum (around 21–200 kg/m^3) are located at ocean ridges (Atlantic and Pacific ridges). The standard error varies ($C_{\Delta\hat{\rho}}$) between 0 and 270 kg/m^3 with a global average of 44 kg/m^3, with the largest values in ocean ridges. The oceanic ridges are special areas due to the presence of hot spots and light materials very close to the solid-Earth topography, as well as the thinning of the oceanic crust, modifying the normal MDC between the crust and the mantle. Hence, it is natural to obtain large densities from the MDC in those areas. Assuming a normal mantle density value of 3270 kg/m^3, the MDC far from the axis of the ridge would be about 200 kg/m^3 (3200–3000 kg/m^3), and in our study, the estimated MDCs in ocean areas are close to this value. As one moves towards the axis of the ridge, the oceanic crust is increasingly replaced by low-density mantle (partially melted material), with the density values ranging from about 3050–3200 kg/m^3. This case is just over the ridge, where the oceanic crust disappears, and thus the crust-mantle boundary is cropping out (appearing or being exposed at the solid surface of the Earth) at the ocean bottom. Thus, in such an area, it is not meaningful to determine the crust-mantle density contrast, because the oceanic crust has disappeared.

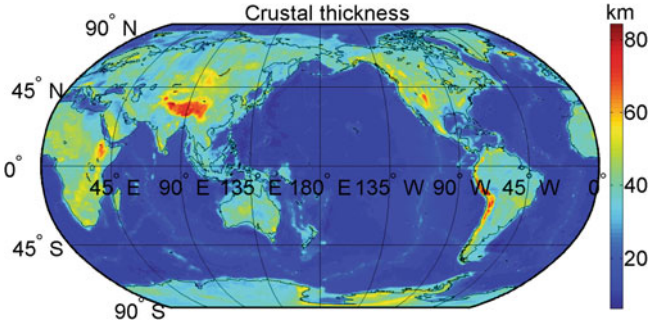

Fig. 8.28 Estimated Moho depth by combined model. Unit: km

For evaluating the estimated MDC, the mean values derived by CRUST1.0, CRUST2.0 (Laske et al. 2000) and PREM (Dziewonski and Andersson 1981) models are used. According to these comparisons, one can see that the global average of 321 kg/m^3 estimated by the least-squares adjustment is 3% smaller than the 330 kg/m^3 derived by the CRUST1.0 model, 20% smaller than the 400 kg/m^3 derived by the CRUST2.0 model, and it is 34% smaller than the 480 kg/m^3 derived by the PREM model. One reason for the large discrepancy between the estimates for $\Delta\rho$ and those by the PREM and CRUST2.0 models could be due to the improvements in the sediments layers in the CRUST1.0 model with respect to the CRUST2.0 and the PREM models.

8.5.8 Moho Recovery Using Gravitational Gradient Data

Over the years, various methods for estimating the Moho depth have been proposed, which have been discussed in this chapter. In this section, the purpose is to use gravitational gradient data for estimating the Moho depth instead of the gravity anomaly/disturbance. The major benefit of the gravitational gradient data to surface gravity is its ability to provide precise and uniform global-data coverage. The gravity-gradient data from GOCE mission were used for the regional and global Moho recovery, for instance, by Sampietro (2009), Bagherbandi (2011), Reguzzoni and Sampietro (2012), Reguzzoni et al. (2013) and Sampietro et al. (2013). Sampietro (2009) studied the problem of recovering the Moho depth from the GOCE data in a simulation study. He considered a local inversion of the satellite gravity-gradiometry (SGG) data by simulating a Moho surface and generating the SGG data based on that. Planar approximations were used in the formulation, and the problems of spatial truncation error of the integral formulas and the behaviour of their kernels were not considered. Sampietro's (ibid) idea is similar to the investigation of Prutkin and Saleh (2009), where they used gravity and magnetic data to invert the Moho discontinuity. Bagherbandi (2011) studied a Moho model obtained from the Vening Meinesz-Moritz (VMM) model and the simulated gravitational gradient data by EGM2008 (Pavlis et al. 2008) in the presence of a white noise of 10 mE (1 mE = 0.0001 mGal/km). The overall goal of this section is to study and evaluate the possibility of using gravitational-gradient data directly for estimating the Moho depth. As follows from the overview of isostatic schemes in this chapter, the VMM isostatic model better approximates the reality by assuming a regional/global isostatic mechanism. In this way, the VMM isostatic gravity data provide more realistic representation of the Earth's inner structure than those based on classical isostatic schemes. Following this principle, the VMM isostatic problem is formulated to find the Moho depth from vertical gravitational gradients. To treat the satellite gradiometry data means first to convert the gravitational gradients to gravity disturbances close to the Earth surface, and then correcting the data by using additive gravity corrections (see Sect. 8.5.4).

8.5.8.1 First-Order Moho Depth from Vertical-Vertical Data T_{rr}

From Eq. (8.69), it holds that the Moho depth to first-order approximation is:

$$D_1 = K\sum_n \frac{2n+1}{n+1}\delta g_n^B, \tag{8.109}$$

where $K = -1/(G\Delta\rho)$ and δg_n^B is the Bouguer gravity disturbance.

Considering also Eqs. (3.68a, 3.68b) with T_{rr}^B being the Bouguer disturbance potential gradient:

$$\delta g^B = R\sum_n \frac{(T_{rr}^B)_n}{n+2} = \frac{R}{4\pi}\iint_\sigma \sum_n \frac{2n+1}{n+2}P_n(t)T_{rr}^B d\sigma, \tag{8.110}$$

one arrives at the series solution

$$D_1 = KR\sum_{n=0}^{\infty}\frac{2n+1}{(n+1)(n+2)}\sum_{m=-n}^{n}\left(T_{rr}^B\right)_{nm}Y_{nm}, \tag{8.111}$$

and the integral solution

$$D_1 = \frac{KR}{4\pi}\iint_\sigma \mathrm{M}(\psi)T_{rr}^B d\sigma \tag{8.112}$$

with

$$M(\psi) = \sum_{n=0}^{\infty}\frac{(2n+1)^2}{(n+1)(n+2)}P_n(t) = \sum_{n=0}^{\infty}\left[4 + \frac{1}{n+1} - \frac{9}{n+2}\right]P_n(t), \tag{8.113}$$

or (see Eqs. 8.116a–d for s=1)

$$M(\psi) = \frac{2\sqrt{2}}{\sqrt{1-t}} + 9\left(1 - \sqrt{2(1-t)}\right) + (1-9t)\ln\left[1 + \sqrt{2(1-t)}/(1-t)\right] \tag{8.114}$$

or

$$M(\psi) = 9(1-2\upsilon) + 2\upsilon^{-1} + (1-9t)\ln\left(1+\upsilon^{-1}\right), \tag{8.115}$$

where $\upsilon = \sin(\psi/2)$.

Referring to Bois (1961), the following formulas can be derived for $0 \le s \le 1$ when using the notations $L(x) = \sqrt{(1 - 2xt + x^2)}$ and $L(s) = L$:

$$S_1 = \sum_{n=0}^{\infty} s^n P_n(t) = 1/L \tag{8.116a}$$

$$\begin{aligned} S_2 &= \sum_{n=1}^{\infty} \frac{s^n}{n} P_n(t) = \int_0^s \sum_{n=1}^{\infty} x^{n-1} P_n(t) ds = \int_0^s \left(\frac{1}{xL(x)} - \frac{1}{x} \right) dx \\ &= -[\ln 2(1 - xt + L)]_{x=0}^{s} = -\ln(1 - st + L)/2 \end{aligned} \tag{8.116b}$$

$$\begin{aligned} S_3 &= \sum_{n=0}^{\infty} \frac{s^{n+1}}{n+1} P_n(t) = \int_0^s \sum_{n=0}^{\infty} x^n P_n(t) dx = \int_0^s \frac{dx}{L(x)} = [\ln 2(x - t + L(x))]_{s=0}^{s} \\ &= \ln \frac{s - t + L}{1 - t} \end{aligned} \tag{8.116c}$$

and

$$\begin{aligned} S_4 &= \sum_{n=0}^{\infty} \frac{s^{n+2}}{n+2} P_n(t) = \int_0^s \sum_{n=0}^{\infty} x^{n+1} P_n(t) dx = \int_0^s \frac{x}{L(x)} ds \\ &= [L(\mathrm{x}) + t \ln 2(x - t + L(x))]_{x=0}^{s} = L - 1 + t \ln \frac{s - t + L}{1 - t}. \end{aligned} \tag{8.116d}$$

Setting s=1 these formulas are applied in deriving Eqs. (8.114), (8.115) and (8.122b).

8.5.8.2 First-Order Solution to Moho Depth from Gravity-Gradient Horizontal Data

The gravity disturbance can be determined from the horizontal components of gravity gradients G_{PQ} as follows (see Eqs. 3.77):

$$\delta g_P = \sum_{n=0}^{\infty} \delta g_n(P) = \frac{R}{4\pi} \iint_{\sigma} F_2'(\psi) G_{QP} d\sigma_Q, \tag{8.117a}$$

where

$$F_2(\psi) = \sum_{n=2}^{\infty} \frac{2n+1}{n(n+1)} P_n(t). \tag{8.117b}$$

Here the prime means the derivative w.r.t. ψ, and, as $t = \cos\psi$, it follows that $dt/d\psi = -\sin\psi$. Furthermore:

$$G_{PQ} = (T_{zx})_Q \cos\alpha' + (T_{zy})_Q \sin\alpha', \tag{8.119}$$

α' being the azimuth from Q to P. Using Eq. (8.117a) for the Bouguer gravity disturbance and inserting it into Eq. (8.109) lead to the first-order solution to Moho depth of:

$$D_1 = \frac{R}{4\pi G\Delta\rho} \iint_{\sigma} F'(\psi) G^B_{QP} d\sigma_Q, \tag{8.120a}$$

where the kernel function becomes:

$$F(\psi) = \sum_{n=2}^{\infty} \frac{(2n+1)^2}{n(n+1)^2} P_n(t). \tag{8.120b}$$

From Abramowitz and Stegun (1964, p. 334) and Sjöberg (1975, p. 23) it follows that:

$$\frac{d}{d\psi} P_n(t) = \frac{1}{\sqrt{1-t^2}} \frac{n(n+1)}{2n+1} [P_{n+1}(t) - P_{n-1}(t)], \tag{8.121}$$

so that, after a few manipulations, the kernel function in Eq. (8.120a) can be written as:

$$F'(\psi) = \frac{1}{\sqrt{1-t^2}} \left[-\left(t + \frac{3}{2} P_2(t) \right) + \sum_{n=1}^{\infty} \left(\frac{1}{n+2} - \frac{1}{n} \right) P_n(t) \right], \tag{8.122a}$$

or, in a closed form when considering Eqs. (8.116b and 8.116d),

$$F'(\psi) = \frac{1}{\sqrt{1-t^2}} \left[-\left(\frac{3}{2} + t + \frac{3}{2} P_2(t) \right) + 2\upsilon + t \ln\left(1 + \upsilon^{-1}\right) - \ln(\upsilon + \upsilon^2) \right]. \tag{8.122b}$$

where $\upsilon = \sin(\psi/2)$.

In order to see the ability of gravitational gradients to recover Moho geometry, the following global numerical study is presented. The values of the vertical gravitational gradient were generated using the GOCO-03S coefficients

(Mayer-Gürr et al. 2012), complete to a spherical harmonic degree of 180. This spectral resolution is compatible with the 1 × 1 arc-deg data resolution of the CRUST1.0 global-crustal model (Laske et al. 2013). The topographic, bathymetry, ice, sediment and non-isostatic corrections (forward modelling) to gravitational gradients were computed according to Sects. 3.3.8 and 8.5.4. Mathematical expressions for the second-order partial derivatives of the gravitational potential depend on the frame of their definitions. The results of forward modelling are presented in Fig. 8.29, and their statistics are summarized in Tables 8.6 and 8.7.

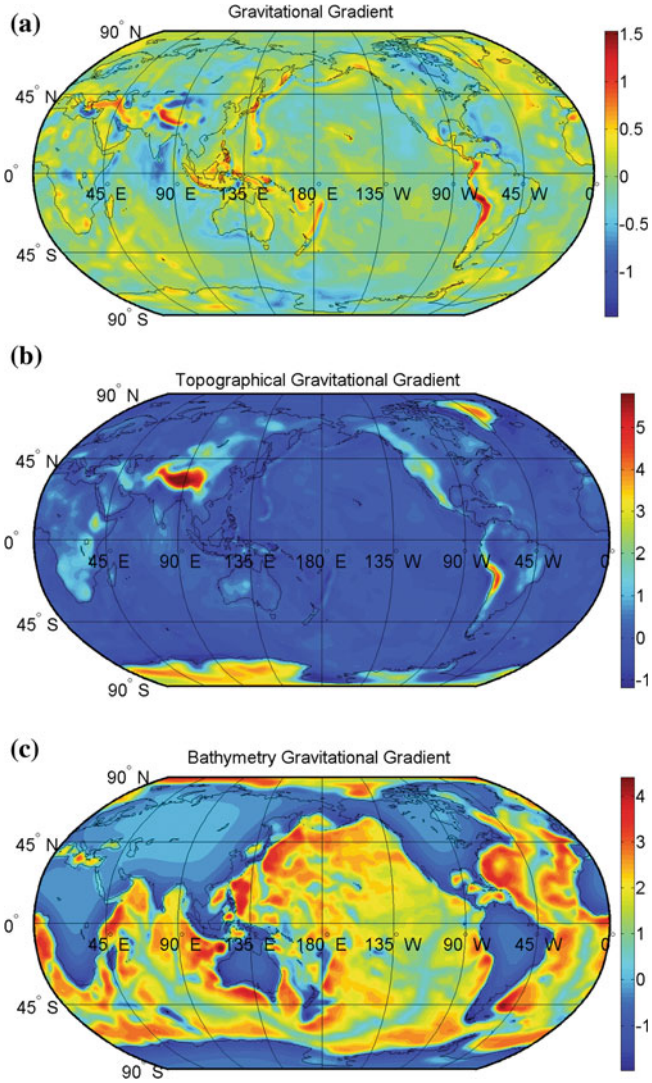

Fig. 8.29 **a** Disturbing gravitational gradient T_{rr} (vertical-vertical component), **b** topography, **c** bathymetry, **d** ice, **e** sediments, **f** non-isostatic effects (NIE) gravitational gradients and **g** disturbing gravitational gradient corrected for the effects of topography, bathymetry, ice, sediment and non-isostatic effect (tbisn). Unit: Eötvös

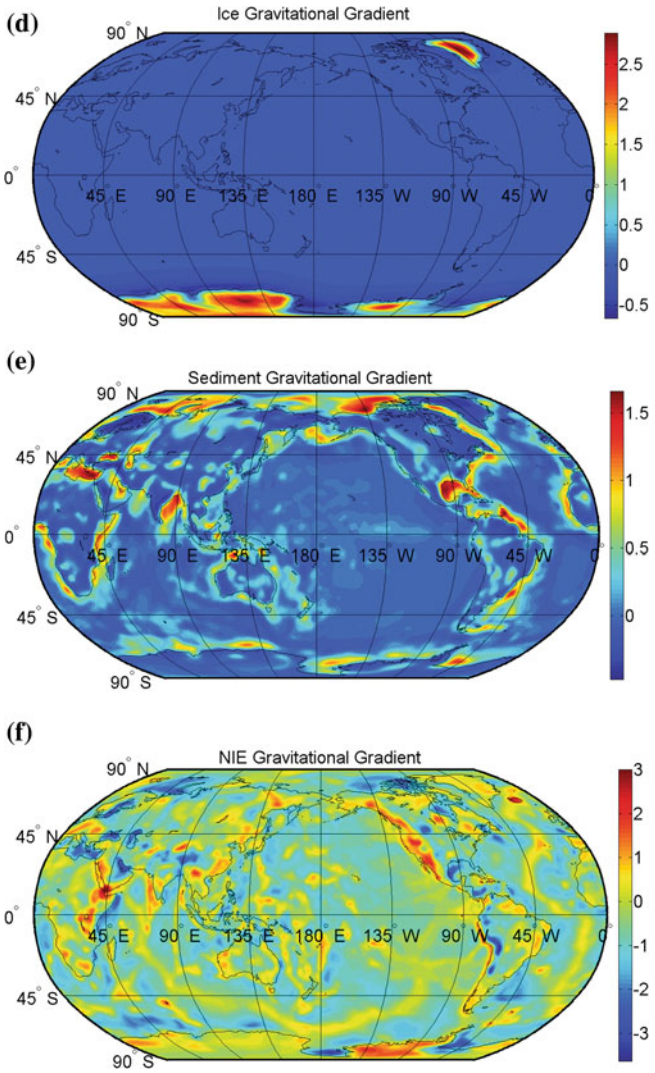

Fig. 8.29 (continued)

According to Fig. (8.29a), the disturbing gravitational gradient T_{rr}, see also Eq. (3.68b), corresponds: (a) to all external masses above the reference ellipsoid GRS80 and (b) to all mass density anomalies within the reference ellipsoid taken relatively to the adopted constant (mean) mass density (so-called mass-density contrasts). Gravitational gradients generated by the homogenous topographic masses are shown in Fig. (8.29b) and by the continental ice masses as presented in Fig. (8.29d). Gravitational gradients generated the mass-density contrasts within bathymetry as plotted in Fig. (8.29c), and, within the sediment layers and the NIEs,

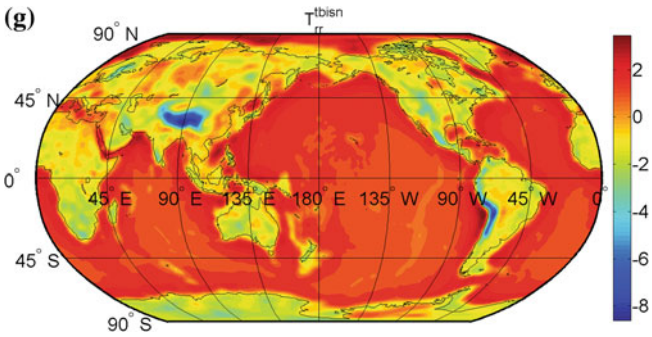

Fig. 8.29 (continued)

Table 8.6 Statistics of gravitational gradient T_{rr}, the topography-corrected gravitational gradient T_{rr}^{t}, the topography-corrected and bathymetry-stripped gravitational gradient T_{rr}^{tb}, the topography-corrected and bathymetry- and ice-stripped gravitational gradient T_{rr}^{tbi}, the topography-corrected and bathymetry- and ice- and sediments-stripped gravitational gradient T_{rr}^{tbis}, the topography-corrected and bathymetry- and ice- sediments and non-isostatic effect (NIE) stripped gravitational gradient T_{rr}^{tbisn}. STD = standard devaition

Gravitational gradients	Max.	Mean	Min.	STD
T_{rr}	1.6	0.0	−1.5	0.24
T_{rr}^{t}	1.1	−0.3	−6.0	0.97
T_{rr}^{tb}	4.6	0.6	−6.1	1.95
T_{rr}^{tbi}	4.6	0.7	−6.1	1.67
T_{rr}^{tbis}	4.7	0.8	−6.2	1.73
T_{rr}^{tbisn}	3.9	0.5	−8.6	1.52

Unit: Eötvös

Table 8.7 Statistics of gravitational gradients generated by topography, bathymetry, ice, sediments and non-isostatic effects (NIE) (vertical-vertical component)

Gravitational gradients	Max.	Mean	Min.	STD
Topography	1.2	−0.3	−6.0	0.98
Bathymetry	4.6	0.9	−2.0	1.37
Ice	3.0	0.1	−0.7	0.51
Sediment	1.7	0.14	0.5	0.31
NIE	3.2	−0.35	−3.6	0.71

Unit: Eötvös

are shown in Fig. (8.29e–f). Numerical values of all gravitational gradients are given in Eötvös= 10^{-9} s^{-2}. For example, the large negative values show subduction zones and the large positive ones are under orogen areas in Tibet.

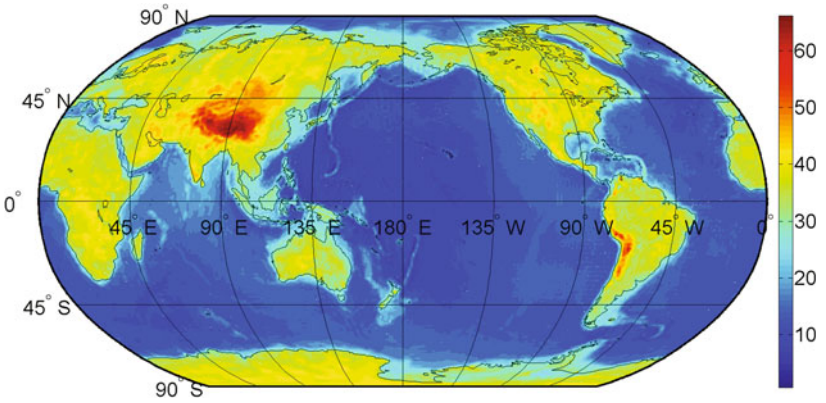

Fig. 8.30 Moho depth obtained from the disturbing gravitational gradient T_{rr}. Unit: km

Table 8.8 Statistics of the VMM–Moho solution (shown in Fig. 8.30) computed using gravitational gradients

Quantities	Max.	Mean	Min.	STD	RMS
Moho depth using gravity gradients (D_1)	66.5	23.9	2.0	11.87	
CRUST1.0	74.8	22.9	7.4	12.4	
$D_1 - CRUST1.0$	20.7	−1.0	−16.9	3.7	3.8

Unit: km

The vertical gradient of the crust and non-isostatic stripped disturbing gravitational gradients (T_{rr}^{tbisn}) varies globally between −8.6 E and 3.9 E (Table 8.6) at the altitude of 250 km above Earth's surface (satellite elevation). These data were used to compute the global Moho solutions based on solving the VMM problem by Eq. (8.111). The global Moho map is shown in Fig. 8.30, and statistics of the results are summarized in Table 8.8. The maximum Moho deepening is under the orogens of the Himalaya, the Andes and Tibet, while the minimum Moho depths are beneath mostly the oceanic areas. When checking the computed results, it is clear that the largest spatial variations of these refined gradient data (Fig. 8.29g) closely agree with the largest Moho variations (Fig. 8.30) along continental margins and geological boundaries between large orogens and continental basins. Over oceans, the locations of hotspots and mid oceanic ridges can also be recognized on the map of gravity gradient. To validate gravimetric results we compare with CRUST1.0 seismic model (Laske et al. 2013). The comparison reveals that the obtained Moho depth is systematically biased with respect to the CRUST1.0 model, implying that the mean Moho depth difference ($D_1 - CRUST1.0$) is −1 km, and the RMS difference is 3.8 km.

8.6 Tectonic Stress in the Mantle

The link between tectonics and mantle convection is one of the oldest and most challenging problems in the history of geodynamics. According to Runcorn (1967), flow in the Earth's mantle is inferred from continental drift. The original theories of mantle convection were put forward by Holmes (1931). He defined a hypothesis, which shows the existence of mantle convection. The physics behind the hypothesis of continental drift as articulated by Wegener (1912, 1924) was explained in the 1960s under the new name plate tectonics as driven by mantle convection, that gives rise to phenomena like sea-floor spreading at mid-ocean ridges, deep-sea trenches and transform faults (e.g. McKenzie and Parker 1967; Le Pichon 1968). The role of plate rheology is important to understand lithospheric stress caused by mantle convection. Tectonics is related to the processes that control the structure and properties of the Earth's crust. Mantle convection is the fundamental engine in Earth's dynamics. The relationship between tectonics and convection in the mantle is still debated, but new information from seismic tomography and modelling capacities promises to provide more insight in the near future.

When hot material from the Earth's interior rises, cold material sinks and the induced flow governs plate tectonics and volcanic activity. Earth and other planets, such as Venus and Mars, retain some heat at depth from their formation, so that their surfaces are cold relative to their hotter interiors (Bercovici 2010). The theory of continental drift has been developed and completed to plate tectonics in the 1960s, e.g. by Runcorn (1962a, b). Convection is also a fundamentally important process that controls much of the dynamics of the Earth's atmospheric, oceanic, mantle and outer-core systems. The study of the theory of mantle convection is important for many applications, e.g. plate velocities, sea-floor subsidence, volcanism, gravity change, stress in the mantle, glacial isostatic adjustment, etc. Figure 8.31 shows the Earth's structure (crust, mantle and outer core) and mantle convection. Upwelling melted- mantle material reaching the Earth's surface undergoes and this melt reaches the surface in various types of volcanic settings: at mid-ocean ridges, where tectonic plates spread apart and draw mantle up into the opening gap, at plate boundaries and ocean-islands or hotspots, which are

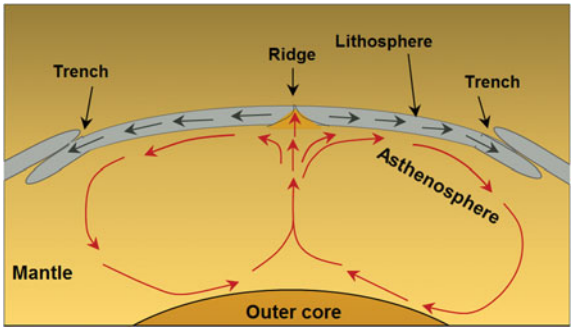

Fig. 8.31 The Earth's structure and mantle convection (provenance unknown)

anomalously productive and localized volcanic features and stresses, not necessarily associated with tectonic activity (Hawaii, Iceland and the East-African Rift being the most conspicuous examples of such a feature).

In order to gain a better understanding of crustal and lithospheric deformation and seismicity, stress modelling in the lithosphere caused by mantle convection is necessary (Becker and O'Connell 2001). The property of Earth's rheology called viscosity, i.e. resistance to flow, is the most crucial parameter to model tectonics and stresses in the mantle (see Sect. 8.8). There are various types of viscosity defined in slightly different ways (Davies 1999). The mantle-flow distribution may be obtained from the pattern of zones of compression and tension in the Earth's crust and also from the low harmonics of the geopotential (Runcorn 1967).

8.6.1 Stress

When tectonic plates move toward each other, they exert a force on the body mass, which is called stress. Tectonic stress is defined as the amount of force per unit area that acts on a mass as the Earth's lithosphere moves. The force, or stress, across a plate depends on the direction of motion of the plate. Mathematically, Cauchy's stress tensor on a surface element of the mass is defined by the stress vector

$$\mathbf{t} = \lim_{\delta P \to 0} \frac{\delta \mathbf{F_n}}{\delta P}, \tag{8.123}$$

where $\boldsymbol{n}$ denotes the normal to the surface element force $\delta \mathbf{F}$ acting on a surface element δP (see Fig. 8.32).

Cauchy's law explains that there is a stress tensor $\mathbf{S}$, which maps $\boldsymbol{n}$ to a surface P to the traction vector acting on that surface:

$$\mathbf{t} = \mathbf{S}\,\mathbf{n} \tag{8.124a}$$

Hence, the components of full stress vector can be written:

$$\begin{aligned} t_1 &= \sigma_{11} n_1 + \sigma_{12} n_2 + \sigma_{13} n_3 \\ t_2 &= \sigma_{21} n_1 + \sigma_{22} n_2 + \sigma_{23} n_3 \\ t_3 &= \sigma_{31} n_1 + \sigma_{32} n_2 + \sigma_{33} n_3, \end{aligned} \tag{8.124b}$$

where:

$$\mathbf{S} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}, \tag{8.124c}$$

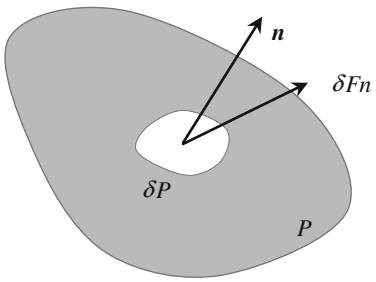

Fig. 8.32 Stress acting on a small part of a body mass

and

$$n^T = [\, n_1 \quad n_2 \quad n_3 \,]. \tag{8.124d}$$

The following characteristics can be mentioned for the stress tensor:

- Nine components of which six are independent (because $\sigma_{ij} = \sigma_{ji}$),
- Stress values which depend on orientation relative to a set of reference axes,
- Six of the nine components become zero in a particular orientation,
- The diagonal components are the normal stresses acting on a surface element normal to the coordinate axes, while the other components represent shear stresses acting on the surface elements.

The stress components could be written in a Cartesian coordinate system using three arbitrary base vectors $\vec{i}, \vec{j}$ and $\vec{k}$ as:

$$\begin{aligned} t_x &= \sigma_{xx}\vec{i} + \sigma_{xy}\vec{j} + \sigma_{xz}\vec{k} \\ t_y &= \sigma_{yx}\vec{i} + \sigma_{yy}\vec{j} + \sigma_{yz}\vec{k} \\ t_z &= \sigma_{zx}\vec{i} + \sigma_{zy}\vec{j} + \sigma_{zz}\vec{k}. \end{aligned} \tag{8.125}$$

The stress vector and its components are expressed in units of Pascal (Pa) = N/m^2.

8.6.2 *Different Kinds of Stress*

The acting force on a surface could be uniform, implying that the force acts equally in all directions, or differential, implying stress that acts with different magnitudes in different directions. In geoscience, there are usually three types of stress: compression, tension and shear stress (Fig. 8.33):

- Compression is the type of stress that squeezes and compacts a body of rock. It commonly reduces the amount of space that rock occupies, and pushes rocks higher up or deeper down into the crust. Compressive stress happens at

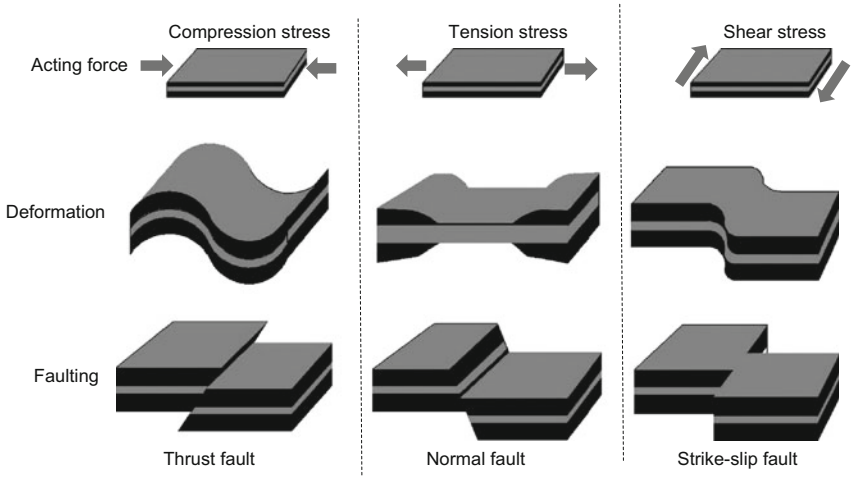

Fig. 8.33 Different types of stress acting on a surface

convergent plate boundaries, and it can make a mountain (e.g. Rocky Mountains in Canada) (Fig. 8.33 left). Thrust faults appear where large blocks of land can be forced over the top of another block of land at a low angle.

- Tension is stress that stretches and pulls a body of rock apart. When rocks are pulled apart by tension, the rocks tend to become thinner. Tension occurs at or near divergent plate boundaries, and within a continent they can form continental rifts and ocean ridge in oceanic crust. (Fig. 8.33 middle). In non-oceanic regions, the fault created by tension stress is called a normal fault.
- Shear stress is the stress component parallel to a given surface, such as a fault plane, that results from forces applied parallel to the surface or from remote forces transmitted through the surrounding rock. Shear stress is experienced at transform boundaries and lithosphere neither is created nor destroyed. In this case, no major topographic effect and volcanic activity appear (Fig. 8.33 right). Shear stress produces the strike-slip phenomenon, simply a strike-slip fault, which can occur while two objects are sliding over each other.
- Principal stresses: The actual values of the six stress components in the stress matrix, **S**, for a given body subjected to loading, will depend on the orientation of the body in the body itself. If the body rotates, it should be possible to find the directions in which the normal stress components take on maximum and minimum values. It is found that, in these directions, the shear components on all faces of the body become zero, and the stress tensor becomes:

$$\mathbf{S} = \begin{bmatrix} \sigma_x & 0 & 0 \\ 0 & \sigma_y & 0 \\ 0 & 0 & \sigma_z \end{bmatrix}. \tag{8.126a}$$

If one of the eigenvalues of the stress tensor is zero, the stress tensor will be a plane stress. That is, there is a Cartesian coordinate system in which the stress tensor has the form

$$\mathbf{S} = \begin{bmatrix} \sigma_x & 0 & 0 \\ 0 & \sigma_y & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{8.126b}$$

- Like the normal stress, the shear stress will also have a maximum in a certain direction. This direction can be determined by taking a derivative of the shear-stress rotation equation with respect to the angle and equate it to zero. The maximum and minimum shear stress can be calculated by:

$$\tau_{\max} = \sqrt{\frac{\left(\sigma_x - \sigma_y\right)^2}{2} + \sigma_{xy}^2}, \tag{8.127}$$

where σ_x is normal stress in x direction, σ_y is normal stress in y direction and σ_{xy} is shear stress perpendicular to x axis and in y direction. The minimum shear stress will be the same absolute value as the maximum, but in the opposite direction.

- Stress invariants: when the stress tensor is expressed with reference to sets of axes oriented in different directions, the components of the tensor change. However, certain functions of the components do not change. These are known as stress invariants, expressed as I_1, I_2 and I_3, where:

$$\begin{aligned} I_1 &= \sigma_{xx} + \sigma_{yy} + \sigma_{zz} \\ I_2 &= \sigma_{xx}\sigma_{yy} + \sigma_{yy}\sigma_{zz} + \sigma_{zz}\sigma_{xx} - \sigma_{xy}^2 - \sigma_{yz}^2 - \sigma_{zx}^2 \\ I_3 &= \sigma_{xx}\sigma_{yy}\sigma_{zz} + 2\sigma_{xy}\sigma_{yz}\sigma_{zx} - \sigma_{xx}\sigma_{yz}^2 - \sigma_{yy}\sigma_{zx}^2 - \sigma_{zz}\sigma_{xy}^2, \end{aligned} \tag{8.128}$$

As has been mentioned before, plate tectonics is a clear signature of mantle convection. The locations of major plate boundaries are shown in Fig. 8.34, which illustrates that the plates move and interact differently at the boundaries. The movements at the boundaries can be detected by studying the stress components. The deformation of an object is referred to as stress, and stress takes many forms. Several important features are formed at the boundaries, such as deep-sea trenches, volcanoes, earthquakes, hot spots, etc. The most significant earthquakes occur at the plate boundaries with the maximum stresses (Fig. 8.35).

The trenches of Peru-Chile, Mariana and Tonga are the most important regions for divergent plate boundaries. Hot spots are points on the Earth's surface located directly over mantle plumes with magma rising from deep within the mantle. Over time this process creates island chains in the seas and oceans. For instance, the

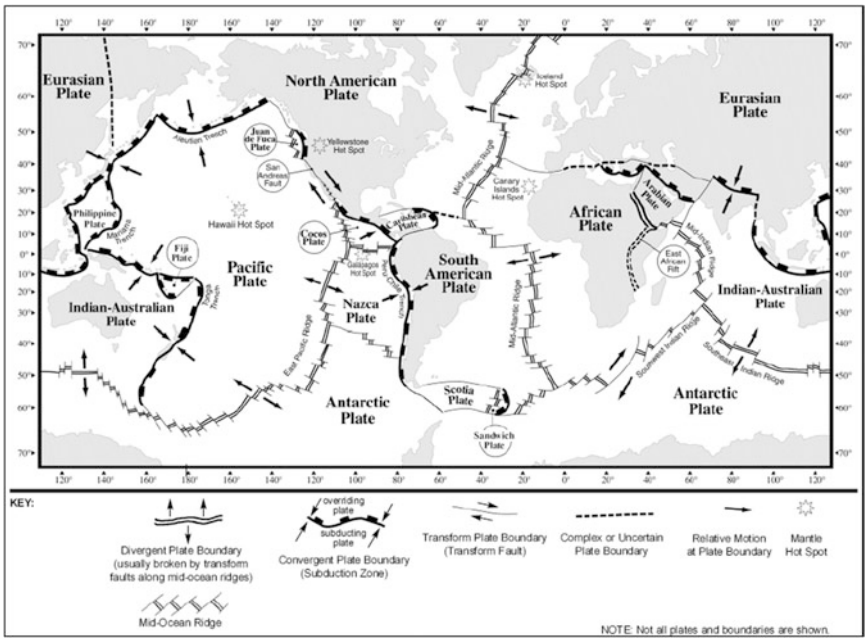

Fig. 8.34 Global tectonic-plates map (provenance unknown) https://ideagirlseverestormpredictionswarnings.wordpress.com/tectonic-maps/

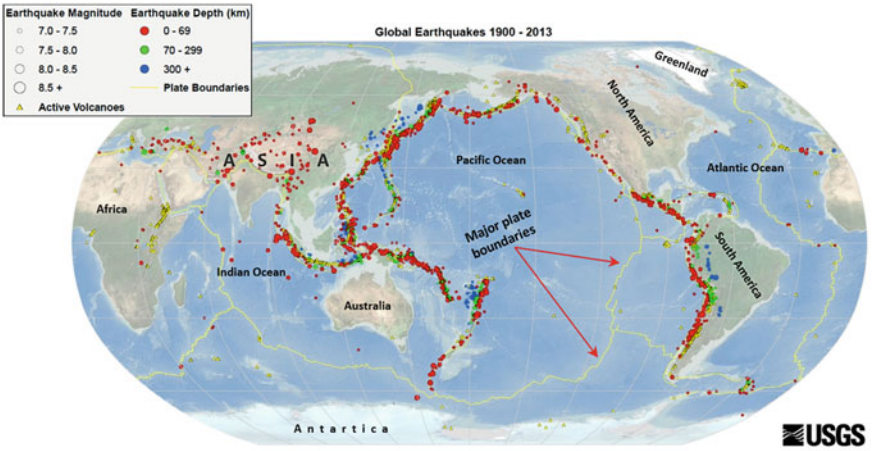

Fig. 8.35 Significant global earthquake epicentre 1900–2013. (http://earthquake.usgs.gov/earthquakes/world/seismicity_maps, 2015-04-05). The epicentre is the point on the earth's surface vertically above the hypocenter (or focus) point in the crust, where a seismic rupture begins

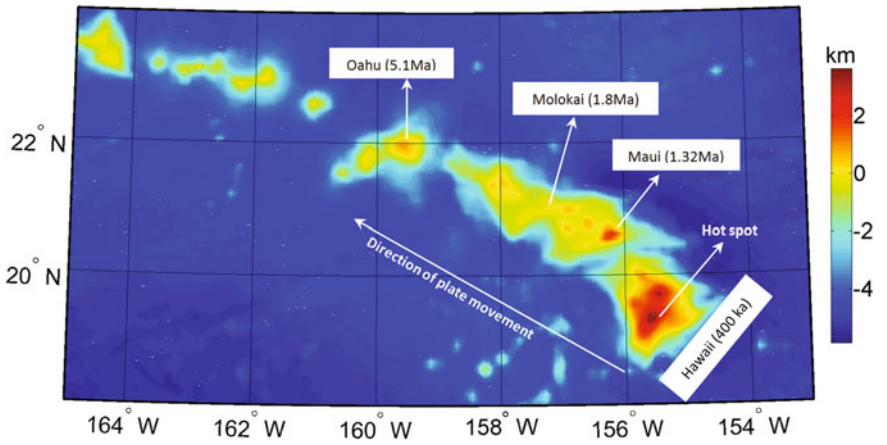

Fig. 8.36 Hot spots in the Hawaiian island chain (background map shows the solid-Earth topography using DTM2006)

Hawaiian island chain located above a hot plume in the Pacific Plate (see Fig. 8.36) is one of important areas to study stress and earthquakes. There is a famous triple junction of three diverging plates in East Africa (see Fig. 8.34).

8.6.3 Determining Stress Using Geometric–Geodesy Techniques

Repetition of geodetic surveys over time offers a great tool to define the pattern of present day relative plate tectonic motions in terms of direction, as well as amount of displacement. Traditionally, these surveys were performed with trilateration methods using electronic distance measurements and angle observations using theodolites (Ortlieb et al. 1989), but more recently GNSS (Scherneck et al. 2010; Riguzzi et al. 2013), Very Long Baseline Interferometry (VLBI) (Hass et al. 2002), Satellite Laser Ranging (SLR) (Noomen et al. 1996; Nieuwland 2003, p. 101) and InSar data (Bürgmann et al. 2000; Hanssen 2001) are commonly utilized.

8.6.3.1 Global Navigation Satellite Systems (GNSS)

The recent development of GNSS networks makes it possible to study (by spatial and temporal resolution) ongoing crustal deformation. GNSS geodetic measurements devoted to active tectonics studies and deformation have been extensively carried out since the early 2000s. For example, the BIFROST (Baseline Inferences

for Fennoscandian Rebound Observations Sea Level and Tectonics) project started in 1993. The purpose was to establish a new 3-D measurement of the movements in the Earth crust using GNSS observations and able to constrain models of the GIA (Glacial Isostatic Adjustment) process in Fennoscandia. Updated station velocities can be found in Lidberg et al. (2007), based on data from the period 1996 to mid-2004. Scherneck et al. (2010) investigated and developed a method for the analysis of crustal strain, determined by station networks that are continuously measured by GNSS. The data they analysed are eight-years' worth of daily results from continuous BIFROST GPS measurements in the permanent networks of the Nordic countries and their neighbours. Reducing the observations with best-fitting predictions for the effects of GIA, they found maximum strain rates predominately in the interior of the rebound area. Riguzzi et al. (2013) also used GPS data to study strain rates in active faults in Italy. They found that geodetic strain rates, integrated with the knowledge of active faults, indicates that the most hazardous seismic areas are those with lower strain rates, where active faults are possibly approaching the end of the seismic cycle (see also Devoti et al. 2014).

8.6.3.2 Very Long Baseline Interferometry (VLBI)

VLBI is a type of astronomical interferometry used in radio astronomy, where a signal from an astronomical radio source, such as a quasar, is collected at multiple radio telescopes on Earth. The distance between the radio telescopes is then calculated using the time difference between the arrivals of the radio signal at the telescopes. This enables observations of an object that are made simultaneously by many radio telescopes to be combined, emulating a telescope with a size equal to the maximum separation between the telescopes.

Today, baseline measurements between geodetic VLBI stations are achieved with an accuracy of better than 2 mm. Since 1990, the European fixed-station geodetic VLBI network has been observing on a regular basis in order to determine surface motion in Europe. The main purpose of establishing VLBI systems is the determination of surface deformation (strain rate) in Europe, and the other goal is to provide a stable reference network for other geodetic techniques used in the area, e.g. GPS networks (Haas et al. 2002).

8.6.3.3 Satellite Laser Ranging (SLR)

SLR is a technique that measures the round-trip time of flight of ultra-short pulses of light to satellites equipped with retroreflectors. This provides instantaneous range measurements at millimetre-level precision, which can be accumulated to provide accurate measurement of orbits and a host of important scientific data for tectonic stress in the Earth's crust (Nieuwland 2003, p. 101). Its capability to study the

variations over time in the Earth's gravity field and to monitor changes with respect to the reference frame (e.g. The International Terrestrial Reference Frame (ITRF)), together with the capability to monitor vertical motion, makes it unique for modeling and evaluating long-term deformations and changes. The ability of the SLR system is not only in the determination of plate tectonics, but it is also applicable to provide a reference system for post-glacial rebound, sea level and ice volume changes, determining the temporal mass redistribution of the solid Earth, ocean, and atmosphere system and in monitoring the response of the atmosphere to seasonal variations in solar heating.

8.6.3.4 Interferometric Synthetic Aperture Radar (InSar)

InSar is a radar technique used in geodesy and remote sensing. This geodetic method uses two or more synthetic-aperture-radar (SAR) images to generate maps of surface deformation or digital elevation, using differences in the phase of the waves returning to the satellite (Massonnet and Feigl 1998; Hanssen 2001) or aircraft. The technique can potentially measure millimetre-scale changes in deformation over spans of days to years. It has applications for geophysical monitoring of natural hazards, for example, earthquakes, volcanoes and landslides, and in structural engineering, in particular, monitoring of subsidence and structural stability (Bürgmann et al. 2000). Whereas terrestrial observations, like GNSS, provide sparse point-wise data, the InSar technique can deliver very dense data, but only 1-D observations in the direction of satellite to ground. Many studies that have been published in recent years demonstrate the contributions of InSAR in the Earth sciences, e.g. studies of topography, surface deformation, in the fields of crustal deformation research including surface displacements related to tectonic deformation.

8.6.3.5 Strain Analysis Using Geodetic Observations

In order to study and analyse deformation phenomena in active zones, it's important to determine the following two parameters: (a) displacement vectors from geodetic data and (b) the principal components of the strain tensor from the displacement vector. Generally, the deformation is, in terms of strain (E), defined by:

$$\varepsilon = \frac{\Delta \mathrm{L}}{\mathrm{L}_0} = \frac{\sigma}{E}, \tag{8.129}$$

where the strain (which is unitless) is the deformation (ΔL elongation) per unit of the original length L. The last part of Eq. (8.129) yields the relation between strain and normal stress (σ). ΔL is the change of length, L_0 denotes the initial length (L_0) and E is Young's modulus (the modulus of elasticity).

In order to model the deformation parameters (stress and strain) of a deformable body, let $\mathbf{r} = (x, y, z)$ represent the position of point P_1 at epoch t_0 before deformation and $\mathbf{r}' = (x', y', z')$ denote the position of point P_1 after deformation. Then the displacement vector of P_1 can be written as:

$$\mathbf{d} = \mathbf{r}' - \mathbf{r} = (x' - x \quad y' - y \quad z' - z)^T = (u \quad v \quad w)^T, \tag{8.130}$$

where u, v, and w represent the displacement components in the x, y and z directions, respectively.

Let point P_2 be another material point in the infinitesimal neighbourhood of P_1. Figure 8.37 illustrates the displacement $\mathbf{d}$, which the points in the neighbourhood of P_1 experience if the body is deformed. The relative position vector d$\mathbf{r}$ of the original configuration is deformed into $d\mathbf{r}'$ after deformation.

The change in the neighbourhood is characterized by the 1inear neighbourhood transformation (Fraeijs de Veubeke 1979; Horemuž 2010):

$$d\mathbf{r}' = \nabla \mathbf{r}'^T d\mathbf{r}, \tag{8.131}$$

where ∇ denotes the gradient operator yielding the Jacobian matrix $(\nabla \mathbf{r}'^T)$ of the transformation from position $\mathbf{r}$ to $\mathbf{r}'$:

$$\nabla \mathbf{r}'^T = \begin{bmatrix} \frac{\partial x'}{\partial x} & \frac{\partial x'}{\partial y} & \frac{\partial x'}{\partial z} \\ \frac{\partial y'}{\partial x} & \frac{\partial y'}{\partial y} & \frac{\partial y'}{\partial z} \\ \frac{\partial z'}{\partial x} & \frac{\partial z'}{\partial y} & \frac{\partial z'}{\partial z} \end{bmatrix} = \mathbf{F}, \tag{8.132}$$

and the differential displacement d$\mathbf{d}$ is obtained by:

$$\mathrm{d}\mathbf{d} = \mathrm{d}\mathbf{r}' - \mathrm{d}\mathbf{r} = \nabla \mathbf{r}'^{\mathrm{T}} \mathrm{d}\mathbf{r} - \mathrm{d}\mathbf{r} = (\mathbf{F} - \mathbf{I})\mathrm{d}\mathbf{r} = \mathbf{E}\,\mathrm{d}\mathbf{r}, \tag{8.133}$$

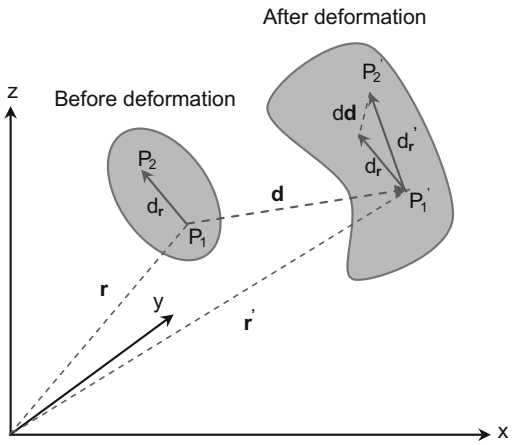

Fig. 8.37 Deformation of a body (cf. Horemuž 2010)

where

$$\mathbf{E} = \mathbf{F} - \mathbf{I} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} & \frac{\partial w}{\partial z} \end{bmatrix} = \begin{bmatrix} e_{xx} & e_{xy} & e_{xz} \\ e_{yx} & e_{yy} & e_{yz} \\ e_{zx} & e_{zy} & e_{zz} \end{bmatrix}. \tag{8.134}$$

The matrix $\mathbf{E}$ is asymmetric, and consequently it can be decomposed as:

$$\mathbf{E} = \frac{1}{2}\left(\mathbf{E} + \mathbf{E}^T\right) + \frac{1}{2}\left(\mathbf{E} - \mathbf{E}^T\right). \tag{8.135}$$

The first term (related to the strain tensor) is a symmetric matrix, and second term (related to rotation of the deformed body) is a skew symmetric matrix. $\mathbf{E}$ can also be written in the form of tensor notation as:

$$\mathbf{E} = \boldsymbol{\varepsilon}_{ij} + \boldsymbol{\omega}_{ij}, \tag{8.136}$$

where $\boldsymbol{\varepsilon}_{ij} = \frac{1}{2}\left(\mathbf{e}_{ij} + \mathbf{e}_{ji}\right)$ and $\boldsymbol{\omega}_{ij} = \frac{1}{2}\left(\mathbf{e}_{ij} - \mathbf{e}_{ji}\right)$ and i and $j = x, y, z$ are the strain tensor and $\boldsymbol{\omega}_{ij}$ represents the rotations of the rigid body, respectively. The components $\varepsilon_{xx}, \varepsilon_{yy}, \varepsilon_{zz}$ are called the normal strains and describe the extension, or change in length of the vectors originally parallel to the coordinate system axes. The components $\varepsilon_{xy}, \varepsilon_{xz}, \varepsilon_{yz}$ are called shear strains and describe the change of angles between the coordinate axes.

8.6.4 *Determining Stress by Disturbing Potential Components*

Large mass distributions and irregularities in the Earth's layers can be studied in components of the Earth's gravity field such as geoid and long-wavelength components of the gravity field. Runcorn (1967) presented the stress between two different Earth layers using the Navier-Stokes equation. He further assumed that the upper part of the Earth consists of two layers, the outer layer, the rigid crust, and the inner viscous layer, the mantle, which satisfy the following conditions:

1. The size of the upper part is large enough so that the gravitational effect of the lower boundary can be neglected.
2. The shape of the upper boundary is such that hydrostatic equilibrium is provided.
3. The coefficient of mantle viscosity is constant.

Assuming that the disturbing potential of the Earth's gravity field is T, the stress components exerted by the mantle convection on the crust in the north-south and east-west directions are given by (Runcorn1967) as:

$$\sigma_x = \frac{Mg}{4\pi(R-D_0)^2}\sum_{n=2}^{\infty}\left(\frac{R}{R-D_0}\right)^{n+1}\frac{2n+1}{n+1}\frac{\partial S_n}{\partial\theta} \tag{8.137a}$$

and

$$\sigma_y = \frac{Mg}{4\pi(R-D_0)^2}\sum_{n=2}^{\infty}\left(\frac{R}{R-D_0}\right)^{n+1}\frac{2n+1}{n+1}\frac{\partial S_n}{\sin\theta\partial\lambda} \tag{8.137b}$$

respectively, where $S_n = T_n R/(GM)$, M is the mass of the Earth, g is mean gravity at the Earth's surface, D_0 is the mean Moho depth and T_n is disturbing potential (Eq. 8.1). In practice the summations in Eqs. (8.137a, b) must be finite, say, $n_{\max}$. By inserting the disturbing potential (Eq. 8.1) into Eqs. (8.137a and b), the stress components become:

$$\sigma_x = \frac{Mg}{4\pi(R-D_0)^2}\sum_{n=2}^{n_{\max}}\left(\frac{R}{R-D_0}\right)^{n+1}\frac{2n+1}{n+1}\sum_{m=-n}^{n}C_{nm}Q_m(\lambda)\frac{\partial\bar{P}_{n|m|}(\theta)}{\partial\theta}, \tag{8.138a}$$

$$\sigma_y = \frac{Mg}{4\pi(R-D_0)^2}\sum_{n=2}^{n_{\max}}\left(\frac{R}{R-D_0}\right)^{n+1}\frac{2n+1}{n+1}\sum_{m=-n}^{n}m\,C_{nm}Q_{-m}(\lambda)\frac{\bar{P}_{n|m|}(\theta)}{\sin\theta}, \tag{8.138b}$$

where

$$Q_m(\lambda) = \begin{cases}\cos m\lambda & m \le 0\\ \sin m\lambda & m > 0\end{cases}, \tag{8.139}$$

and (Ilk 1983, Z.1.44):

$$\frac{\partial\bar{P}_{n|m|}}{\partial\theta} = \frac{1}{2}\left[(n+|m|)(n-|m|+1)P_{n,|m|-1} - P_{n,|m|+1}\right]. \tag{8.140}$$

Equation (8.138b) contains a singular term $1/\sin\theta$ in the first-order derivative of the Legendre polynomial. Some attempts have been performed to remove this singularity (see, for example, Petrovskaya and Vershkov 2006).

Finally, the magnitude and direction (azimuth) of the stress S can be expressed by:

$$\mathrm{S} = \sqrt{\sigma_x^2+\sigma_y^2}, \tag{8.141a}$$

and

$$\alpha = \arctan\left(\frac{\sigma_y}{\sigma_x}\right). \tag{8.141b}$$

8.6.4.1 Stress and Deflection Components

The low-degree components of the deflection of the vertical in the north-south and east-west directions are given by Eq. (8.1) inserted into Eqs. (3.41a) and (b):

$$\begin{aligned}\xi &= -\frac{1}{R}\frac{\partial N}{\partial \varphi} = \frac{1}{R}\frac{\partial N}{\partial \theta} = \frac{GM}{R^2\gamma_0}\sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n} C_{nm}\frac{\partial Y_{nm}(P)}{\partial \theta} \\ &= \sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n}\xi_{nm}Y_{nm}(P) = \sum_{n=2}^{n_{\max}}\xi_n(P),\end{aligned} \tag{8.142a}$$

and

$$\eta = -\frac{GM}{R^2\gamma_0}\sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n} C_{nm}\frac{\partial Y_{nm}(P)}{\sin\theta\partial\lambda} = \sum_{n=2}^{n_{\max}}\sum_{m=-n}^{n}\eta_{nm}Y_{nm}(P) = \sum_{n=2}^{n_{\max}}\eta_n(P). \tag{8.142b}$$

Hence, the stress components of Eq. (8.138a, b) can be expressed in terms of the deflection components as:

$$\sigma_x = -\frac{g\gamma_0}{4\pi G}\sum_{n=2}^{n_{\max}}\left(\frac{R}{R-D_0}\right)^{n+3}\frac{2n+1}{n+1}\xi_n(P), \tag{8.143a}$$

and

$$\sigma_y = -\frac{g\gamma_0}{4\pi G}\sum_{n=2}^{n_{\max}}\left(\frac{R}{R-D_0}\right)^{n+3}\frac{2n+1}{n+1}\eta_n(P). \tag{8.143b}$$

Figure 8.38a shows the stress-field map under the eastern part of the Eurasian plate. The corresponding numerical investigations described next use Eqs. (8.138a, b) and are based on the Earth gravitational model EGM08, and the spherical harmonic terms of the normal gravity field were computed by the parameters of GRS-80 (Moritz 2000). The EGM was limited to the harmonic window between degrees 13 and 25, according to Liu (1977), to filter the Earth's interior and higher degree of topographic gravity signals. [The low-degree harmonics reflect large scale core-mantle contributions, while higher degrees may stem from short-wavelength mantle convection (Richter and Parsons 1975).]

As shown in Fig. 8.34, the tectonic plates are converging in the Eurasian region, implying compressive stress. This can be clearly seen in Fig. 8.38a. Also, the collision of the Indian and Eurasian plates not only raises the Tibetan plateau but also causes deformation. As mentioned, heat flow in the mantle is the driving force for plate tectonics, which causes the stress. This type of study is one way to explain the reasons for the Earth's crust's deformation and other geodynamical phenomena like earthquakes, plate tectonic, etc., and gravity inversion is an inexpensive tool in understanding such phenomena.

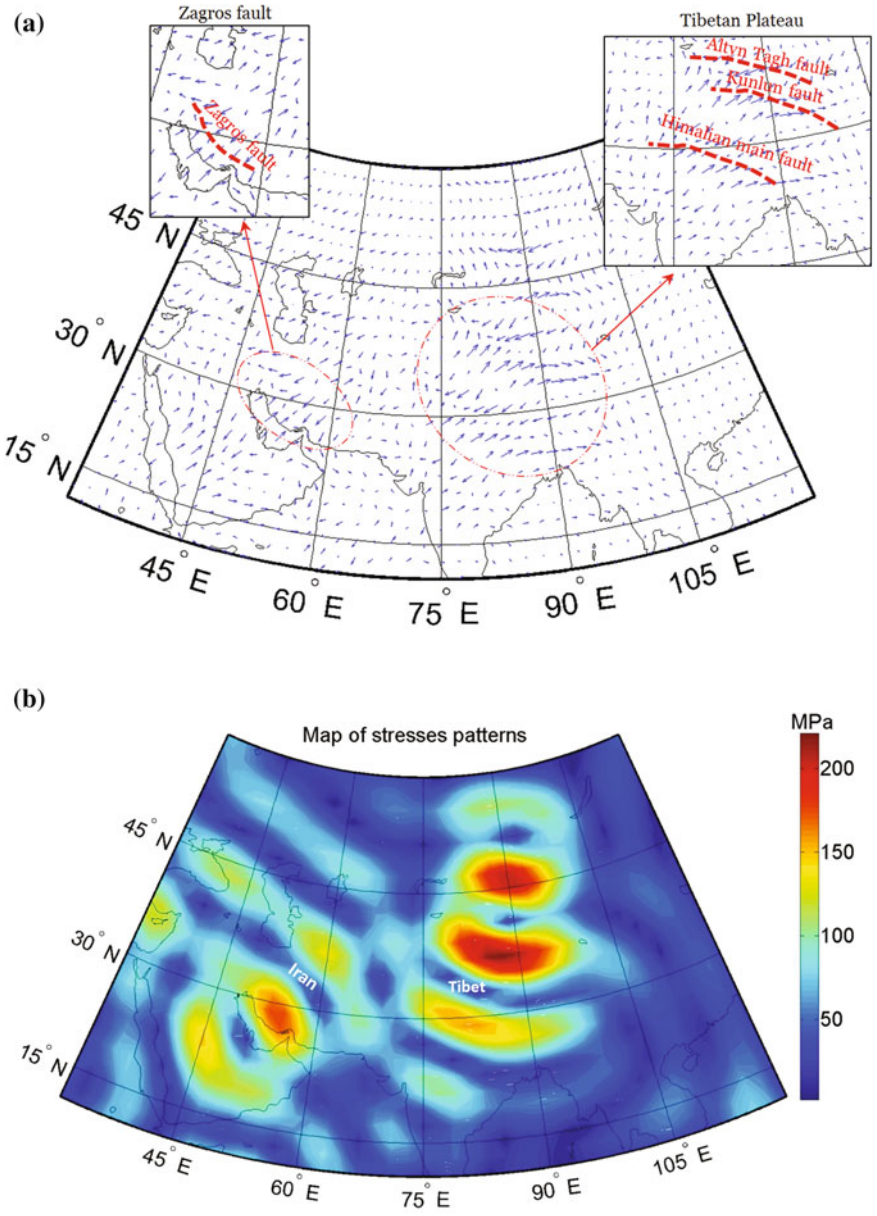

Fig. 8.38 **a** Map of stress patterns (S) beneath Eurasian plate obtained from Eq. (8.116a), **b** Stress exerted by mantle convection under the Eurasian plate using Runcorn (1967) formula on a 1×1 arc-deg grid. Unit: MPa. **c** Stress maps of Tibet (*right*) and Iran (*left*) obtained from the WSM 2008 (Heidbach et al. 2008)

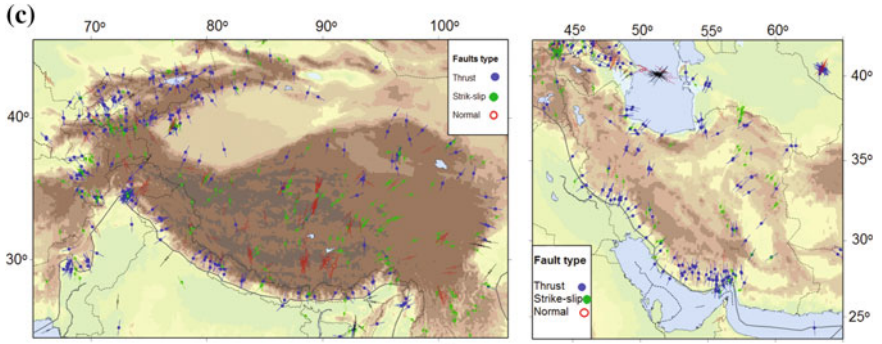

Fig. 8.38 (continued)

Comparisons between stress fields obtained from seismic and gravity data show high correlation. Also, studying the stress fields and lithosphere depths are interesting because of their relationship and high correlation (Pick 1994). The gravimetric results can be compared with the World Stress Map (WSM), presented by the National German Research Centre for Geosciences (GeoForschungsZentrum, GFZ). This map is a global compilation of information on the present-day stress field of the Earth's crust with 21,750 stress data records in its current WSM database release 2008. It is a collaborative project between academia, industry and government that aims at characterizing the stress patterns and to understand the stress sources (see Heidbach et al. 2008 for more information). Different types of stress indicators were used to determine the tectonic stress orientations. They are grouped into four categories: earthquake focal mechanisms, in situ stress measurements, geologic data from fault-slip analysis and volcanic alignments. The comparison of Fig. 8.38a and c shows that the obtained results from gravity inversion are well-matched with the WSM results.

Figure 8.38a shows two active zones in the Eurasian plate. The first one is the Tibetan Plateau that is surrounded by massive mountain ranges and active faults. The second zone is located between the Arabian plate and the Eurasian plate in Iran, i.e. the Zagros Mountain. The major fault zones around the Tibetan Plateau are Altyn Tagh, Kunlun, Haiyuan and Xianshuihe faults. The Altyn Tagh Fault (ATF) is more than 1200 km long and it is a strike-slip fault that forms the northwestern boundary of the Tibetan Plateau. The slip rates have been reported along the majority of the length of the fault and include measurements from geodetic techniques (e.g. GNSS surveys and InSAR). Slip rates determined from GPS campaigns at 90°E are 9 ± 5 mm/year (Bendick et al. 2000), 9 ± 4 mm/year (Wallace et al. 2004) and 11 ± 3 mm/year (Zhang et al. 2007).

The Zagros fold and thrust belt (Zagros FTB) is an approximately 1800 km long zone of deformed crustal rocks, formed in the foreland of the collision between the Arabian and the Eurasian plates. It is the host to one of the world's largest petroleum provinces, containing about 49% of the established hydrocarbon reserves in

fold and thrust belts and about 7% of all reserves globally (Cooper 2007). The Zagros FTB is formed along a section of the plate boundary that is subject to oblique convergence with the Arabian plate moving northwards with respect to the Eurasian plate at a speed of about 30 mm/year. The degree of obliqueness reduces southwards along the Zagros, with the collision becoming nearly orthogonal within the Fars domain. The relative movement between the plates is only partly taken up within the Zagros; the remainder is taken up by deformation in the Alborz mountains and the Lesser Caucasus mountains to the north of the Iranian plateau and along the zone formed by the Greater Caucasus mountains, the Apsheron-Balkan Sill and the Kopet Dag mountains, further north again (Talebian and Jackson 2004).

So far, it is an open question whether Runcorn's gravimetric technique can also be used for more detailed regional studies of stress in the mantle.

8.7 Temporal Changes of the Gravity Field

The Earth's gravity field changes with time due to various geodynamical processes. The motions are secular and periodic. The largest secular changes are caused by mantle convection and glacial isostatic adjustment, as well as plate and intraplate motions. Notable periodic variations are due to Earth's rotation (e.g. the pole tide of period 14 months), seasonal variations caused by variations in atmospheric and hydrologic conditions and tidal variations of various different periods.

Today, most successful data for studying the temporal changes of the gravity field stems from the very long record of Satellite Laser Ranging (Moore et al. 2005) and lately from the accurate, dedicated satellite-gravity mission GRACE (Lemoine et al. 2007). GRACE data greatly complement the Glacial Isostatic Adjustment (GIA) studies because of their long-wavelength characteristic. Therefore, the GRACE sensors are appropriate tools to measure secular gravity changes in the mantle layers. There are different solutions for GRACE monthly, weekly and ten-day solutions provided by several analysis centres. The three main analysis centres are Centre for Space Research (CSR) at the University of Texas at Austin, Jet Propulsion Laboratory (JPL) in Pasadena and the Helmholtz-Zentrum Potsdam, GFZ. In addition, there are solutions from the University of Bonn (ITG), the Centre National d'Etudes Spatiales (CNES) in Toulouse and the Technical University Delft (DEOS Mass Transport model).

8.7.1 Satellite-Based Methods to Study Temporal Variations

By repeating satellite tracks for measuring and determining time-tagged, EGM, linear least squares regression, analysis can be used to determine the temporal

changes of the potential coefficients ($\dot{C}_{nm}$), such that the changes of the external disturbing potential and the gravity anomaly can be determined up to some limited degree M. From Eqs. (8.1) and (8.2a), one obtains the truncated series:

$$\dot{T} = \frac{GM}{R}\sum_{n=0}^{M}\left(\frac{R}{r}\right)^{n+1}\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(\theta,\lambda). \tag{8.144}$$

and

$$\Delta\dot{g} = \frac{GM}{R^2}\sum_{n=0}^{M}\left(\frac{R}{r}\right)^{n+2}(n-1)\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(\theta,\lambda). \tag{8.145}$$

For studying near Earth surface mass changes, it is expedient to use the time derivative of the surface mass density of Eq. (8.5c):

$$\dot{\kappa}(P) \approx \frac{4\pi\bar{\rho}R}{3}\sum_{n=0}^{M}\frac{(2n+1)}{1+k_n}\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(P). \tag{8.146}$$

8.7.2 Temporal Changes of the Geoid

Assuming that the reference radius R is the mean sea level radius, Bruns' formula applied to Eq. (8.144) for $r = R$ yields the temporal change of the geoid height (with $GM/(R\gamma_0) \approx R$)

$$\dot{N} = R\sum_{n=0}^{M}\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(\theta,\lambda). \tag{8.147}$$

Although the geoid partly runs inside the Earth's surface, the topographic effect in Eq. (8.147) is negligible (as long as M is not too large).

From Stokes' formula one obtains also:

$$\dot{N} = \frac{R}{4\pi\gamma_0}\iint_{\sigma} S(\psi)\Delta\dot{g}\,d\sigma, \tag{8.148}$$

and, in the classical approach, $\Delta\dot{g}$ is the time derivative of the free-air gravity anomaly (Sjöberg 1983):

$$\Delta\dot{g} = \dot{g} + \dot{F} = \dot{g} + 0.31\dot{H}, \tag{8.149}$$

where the gravity change is in unit of μGal for $\dot{H}$ in mm/year. In the modern approach, using the temporal change of the surface gravity anomaly, Eq. (8.149) is approximate. However, in both cases, the topographic corrections are negligible.

Finally, we consider also the geoid change in the modified Stokes' formula of Eq. (6.3a), augmented by the temporal changes of the zero- and first-degree harmonics and neglecting additive corrections:

$$\dot{N}_1^{L,M} = \frac{\dot{T}_0 + \dot{T}_1}{\gamma_0} + \frac{R}{4\pi\gamma_0} \iint_{\sigma_0} S^L(\psi)\Delta\dot{g}\,d\sigma + c\sum_{n=2}^{M}\left(Q_n^L + s_n\right)\Delta\dot{g}_n^{EGM}, \tag{8.150a}$$

where $\Delta\dot{g}$ is the same as in Eq. (8.149) and

$$\Delta\dot{g}_n^{EGM} = \frac{GM}{R^2}(n-1)\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(P). \tag{8.150b}$$

Alternatively, Eq. (6.2a) is used (with negligible direct and indirect effects), resulting in the following formula for the change of the geoid height:

$$\tilde{\dot{N}}_1^{L,M} = \frac{R}{4\pi\gamma_0} \iint_{\sigma_0} S^L(\psi)\Delta\dot{g}^M\,d\sigma + \sum_{n=0}^{M}\dot{N}_n^{EGM}, \tag{8.151a}$$

where

$$\dot{N}_n^{EGM} = R\sum_{m=-n}^{n}\dot{C}_{nm}Y_{nm}(P), \tag{8.151b}$$

and the residual anomaly becomes:

$$\Delta\dot{g}_Q^M = \Delta\dot{g}_Q - \frac{GM}{R^2}\sum_{n=2}^{M}(n-1)\dot{C}_{nm}Y_{nm}(Q). \tag{8.151c}$$

Equation (8.147) is suitable for global studies of the change of the geoid, while detailed studies are better determined by Eqs. (8.150a) and (8.151a).

8.7.2.1 Global Scale of Temporal Changes in the Gravity Field

Long-term monitoring of temporal geoid change is important for updating the static geoid model. The change of the geoid with time is caused by the redistribution of masses within the Earth. Generally it can be observed by repeated geodetic observations such as gravimetric observations, including space gravity and gradiometry, and from geodetic height observed by precise levelling and/or GNSS

observations [see Eq. (8.178) and Fig. 8.51]. Major causes are glacial isostatic adjustment and discharge of ice sheets and caps (which can also be modelled by estimating ice history), continental hydrology and climate variability (with additional information from metrological data), groundwater withdrawal (using also hydrological and metrological measurements), earthquakes (considering also seismic and geodetic networks), volcanic activities (with additional information from seismic networks, geodetic networks) (Jacob et al. 2012). The epoch-wise GRACE gravity field models can be used to assess geoid changes. It is important to mention that the gravity data collected by GRACE require smoothing to reduce the effects of errors present in short-wavelength components. Various methods have been proposed to filter the data (Swenson and Wahr 2006; Kusche 2007; Wouters and Schrama 2007; Klees et al. 2008). For example, isotropic Gaussian (Wahr et al. 1998) and non-isotropic (Han et al. 2005) filters are the most used methods. However, none of these methods satisfactorily remove the correlated errors in the data. The correlated noise in the GRACE data deteriorates the signal-to-noise ratio (SNR), but the regularization methods used in the gravity field solutions from the inter-satellite ranging and orbit determination can be manipulated to increase the SNR and affect other metrics of the signal performance. The effect of the noise is particularly on the higher harmonic degrees of the gravity change solutions. The common stripe pattern in the monthly gravity field is the effect of the noise of the higher degrees. Due to the in-orbit configuration of the GRACE mission (Swenson and Wahr 2006), the sectorial and near-sectorial coefficients are more sensitive to the noise than the tesseral coefficients for the same degree. It means that the noise is not white and makes the resulting regularization weighting matrix to shape as a block-diagonal matrix for instance, the other tesseral coefficients are not needed to be regularized within a reasonable approximation (see also Kusche et al. 2009 supplementary documents of the electronic version, Table 1).

According to Kusche et al. (2009), the major problem that users of monthly GRACE gravity field solutions are faced with is that the data has coloured noise that show up in the provided spherical harmonic coefficients. Basically, by truncating the spherical harmonic series at long wavelengths, where the noise is not yet significant, causes the loss of an unacceptably large portion of the signal. The noise can usually be described as stripping patterns (Fig. 8.39). The reason for this problem is the mission geometry, as GRACE twin-satellites fly in the same near polar orbital plane, and the inter-satellite ranging observable used in gravity modelling transforms into distinct along-track sensitivity. Deficiencies in de-aliasing models yield an anisotropic error that cannot be removed by, e.g. the Gaussian isotropic filter. Instead a decorrelation method should be used in the post-processing of the GRACE data, such as those discussed in Kusche (2007), Klees et al. (2008) and Kusche et al. (2009). Kusche's technique uses an a priori synthetic model of the observation geometry for the decorrelation.

The numerical examples that follow are based on CSR data and use an ensemble of anisotropic filters with the smoothing factor $\alpha = 1 \times 10^{13}$ (Kusche et al. 2009). In Kusche's technique the noises are filtered using a priori information on correlated

noise, for which the information of the calibrated error covariance is needed at least for one epoch. Then the filter itself as well as its measures of the performance change with the geographical location consequently, and cannot be represented like in the case of Gaussian filter simply with a half-wavelength radius.

Figures 8.39 and 8.40 show the geoid change for two consecutive months obtained from GRACE mission before and after the non-isostropic filtering of the data according to Kusche (2007), respectively. In the first figure the north-south strips, stemming from the coloured noise, blur the signal, while in Fig. 8.40 the geoid change is visible.

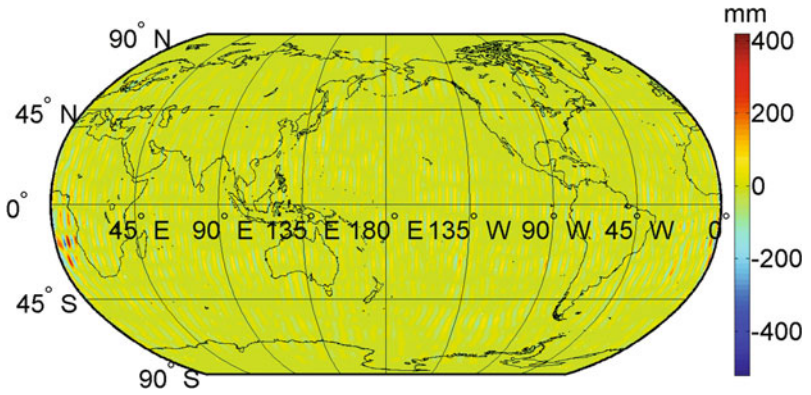

Fig. 8.39 Geoid change using the solution available from CSR (The University of Texas at Austin, Centre for Space Research) processing centre before filtering, Release 05. Time period of data: between April and May 2002. Unit: mm

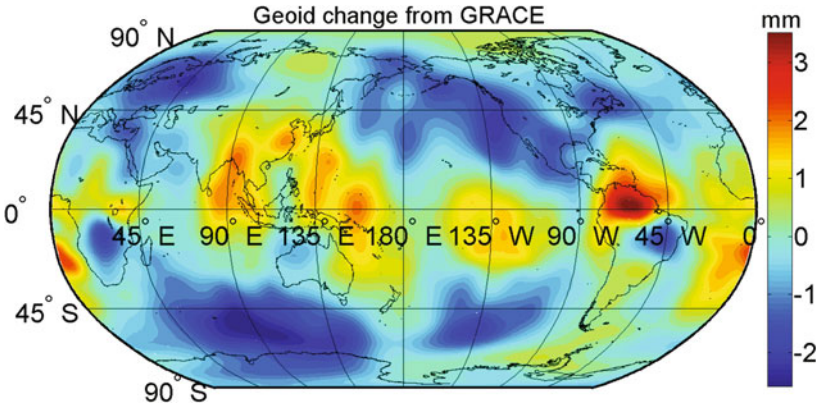

Fig. 8.40 Geoid change using the solution available from CSR using the same data as in Fig. 8.39 but after filtering

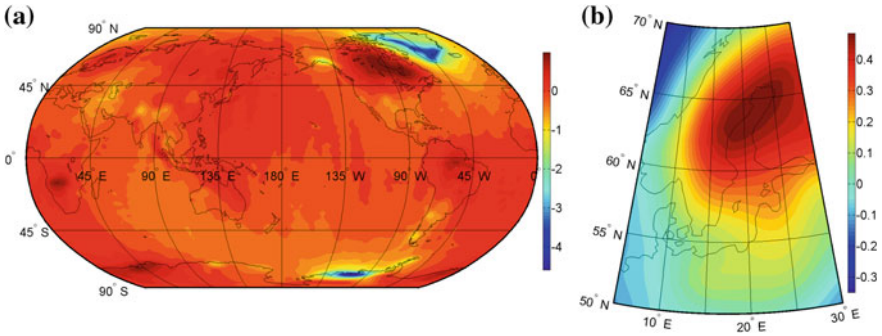

Fig. 8.41 The secular rate of the geoid change ($\hat{b}$, using Eq. 2.6f) by correcting GRACE data using Kusche's anisotropic filter: **a** globally, and **b** regionally in Fennoscandia for the time period of data between April 2002 and October 2014 (about 12.5 years) obtained from linear regression for each point. Unit: mm/year

In order to obtain the secular trend of the geoid change by Eq. (2.6f), a long observation period should be considered. The plot in Fig. 8.41 uses all monthly GRACE data sets between April 2002 and October 2014, computed by DDK4 filter (Kusche et al. 2009). The computations in Fig. 8.41 did not consider any possible long-term periodic variations in the data, and to check for that a special study was performed using Eq. (2.6g) at the location of the maximum uplift rate in Fennoscandia as shown in Table 8.9. The periodical terms considered in this study are the tidal phases observed by the GRACE satellite orbit. According to Ray et al. (2003), the orbit plane of any high-inclination spacecraft like GRACE processes very slowly with respect to the Sun, as the tides generated by the Sun will be aliased into long periods, e.g. semi-annual, annual and longer. These periods are important in climate studies, as they can make an aliased error over long periods. Therefore, the solar tides will affect the GRACE orbit and will be aliased to the long periods. These periods are P1 (171 days), S1 (322 days), S2 (161 days), K1 (7.48 years) and K2 (3.74 years). The results show that among these periods only K1, K2 and S2 are significant. The results show that only the coefficients a, b and c are significant at the risk level of $\alpha = 5\%$, as all other $|T| < t_{\alpha/2,n-1} = 1.96$, where t is the Student's t-distribution statistic. Importantly, the table also shows that the estimated rates of secular change of the geoid height do not differ significantly in the cases of including or not including periodic terms in the analyses.

The secular geoid changes in Fig. 8.41a range between −4.8 and +1.2 mm/year, being primarily due to ice sheet melting and GIA, although these are not the only causes. The largest negative changes of the geoid, caused by ongoing melting and discharging of ice sheets, can be seen in Greenland and W. Antarctica with peaks of −4.2 and −4.8 mm/year, respectively (cf. Sasgen et al. 2005). There are also notable subductions of the geoid in SE Alaska, possibly related with the melting of permafrost, and in the Middle East, in particular at the region of the Caspian Sea, likely due to ground water sinking, as well as in southern S. America, possibly related with glacier retreats.

Table 8.9 Result of regression analysis at the location of the maximum land uplift rate in Fennoscandia with and without periodical terms by Eq. (2.6g) and GRACE data for the time period between April 2002 and October 2014 (about 12.5 years)

Periodic terms	Quantities	Coefficients							
		$\hat{a}$ (mm)	$\hat{b}$ (mm/year)	$\hat{c}$ (mm/year)	$\hat{d}$ (mm/year)	$\hat{e}$ (mm/year)	$\hat{f}$ (mm/year)	$\hat{g}$ (mm/year)	$\hat{h}$ (mm/year)
No periodic terms	Estimated values	−21.09	0.52	–	–	–	–	–	–
	STD	4.04	0.05	–	–	–	–	–	–
	T	−5.22	10.26	–	–	–	–	–	–
K2 = 3.74 years	Estimated values	−20.46	0.51	−6.25	−3.22	–	–	–	–
	STD	4.25	0.05	2.90	3.03	–	–	–	–
	T	−4.81	9.54	−2.15	−1.06	–	–	–	–
K2 = 3.74 years and K1 = 7.48 years	Estimated values	−19.22	0.51	−7.05	−2.17	−4.20	−4.13	–	–
	STD	4.39	0.05	2.99	3.19	3.08	3.18	–	–
	T	−4.37	9.20	−2.35	−0.68	−1.36	−1.30	–	–
K2 = 3.74 years and K1 = 7.48 years and S2 = 161 days	Estimated values	−19.22	0.51	−7.04	−2.16	−4.20	−4.14	0.96	0.11
	STD	4.43	0.05	3.02	3.21	3.10	3.20	2.99	2.99
	T	−4.34	9.13	−2.33	−0.67	−1.35	−1.29	0.32	0.04

$T = \hat{x}/s_{\hat{x}}$, where $\hat{x}$ is the estimated coefficient and $s_{\hat{x}}$ its standard deviation (STD)

As discussed in Sect. 8.8, since the last glacial period the deloading of ice sheets reduces the land masses and increases the GIA induced deformation of the solid Earth in much of northern Europe, Asia, North America, Greenland and Antarctica, which areas were covered by large ice sheets during the last glacial period, and Greenland and Antarctica are still covered by huge ice caps (Figs. 1.4 and 1.5). For example, the ice was as thick as 3 km in Fennoscandia during the last glacial maximum about 20,000 years ago. The deloading of an ice sheet is accompanied by an inflow of subcrustal mass, and this process (GIA) starts directly during the deloading, leading to a mass increase and thereby a geoid uplift. The largest positive geoid changes are seen in NE Canada, centred over Hudson Bay, with a peak of 1.2 mm/year, in Fennoscandia, centred over the Bothnian Bay, with a peak of 0.5 mm/year, and in East Antarctica with 1.1 mm/year, and they are all caused by postglacial rebound (or GIA). The result for Fennoscandia agrees fairly well with previous estimates of Sjöberg (1983) and also with Fig. 8.51, both studies derived from observed land uplift rate data. Sjöberg (1983) estimated that the rebound induced rates of change of surface gravity and geoid height in Fennoscandia have peaks of −1.9 μGal/year and 0.65 mm/year, respectively, the former being consistent with absolute gravity observations, the latter agreeing with the result in Fig. 8.41b. See also Sjöberg and Bagherbandi (2013), where also the eustatic effect is considered. The presented results in North America also agree well with Jacob et al. (2012) based on 97 monthly GRACE solutions from the University of Texas Centre for Space Research Release 04 (spanning September 2002 to October 2010).

There are also regions with relatively large geoid uplifts in SW Africa and in the Amazonas in S. America of unknown origins. The mean global rate of change of the geoid is estimated to -0.06 ± 0.002 mm/year, while over the oceans the mean value is -0.08 ± 0.002 mm/year, the latter being far from the present mean sea level rise of more than +3 mm/year (Nicolls and Casenave 2010). The Intergovernmental Panel on Climate Change (IPGC) reported that the geoid change in Antarctica (cf. Fig. 8.41a) causes a 1.4 mm/year world-wide sea level rise (Bentley and Wahr 1998). A likely reason for this disagreement is that GRACE data badly determines the zero- and first-degree harmonics as well as the second-order (J_{20}) zonal term. Monthly J_{20} coefficients can be taken from long-term satellite laser ranging records (Cheng and Tapley 2004), and the first-degree coefficients, caused by the motion of the geocenter, can be determined by the method of Swenson et al. (2008). These data were not considered in the above plots.

Figure 8.42 shows the Earth's surface-mass changes using GRACE monthly data. Although mass-transport processes and mass anomalies are very complicated in the Earth system, the reasons for the mass changes can be divided into different sources such as ice-mass change and its effect on the sea-level rise, Earth interior motion (mantle dynamics and GIA), ocean dynamics, continental hydrology (e.g. large-scale variation of the continental water-storage changes) and atmospheric effect. Generally, for regions where the geoid goes up, mass increases, and it probably also reflects the changes caused by lower layers of the Earth.

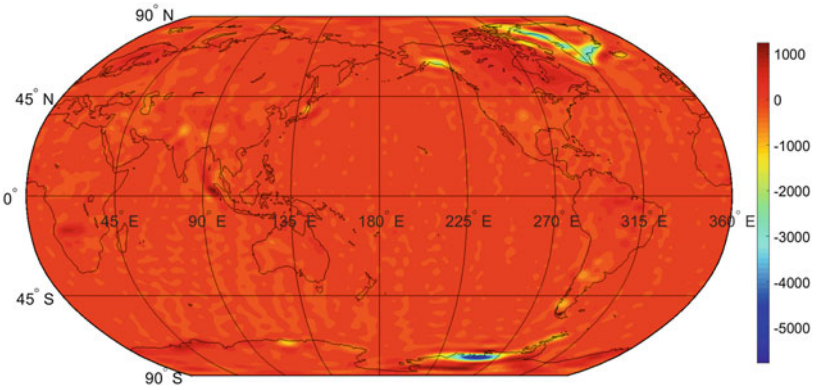

Fig. 8.42 Near Earth surface mass changes ($\dot{\kappa}$) obtained from Eq. (8.146) using GRACE data, computed globally on a 1×1 arc-deg grid of surface points. Time period of data between April 2002 and October 2014. Unit: kg/m^2/year

8.7.2.2 Regional Scale of Temporal Changes in the Gravity Field

It can also be of interest to study geoid and mass changes regionally. We know that degrees up to 4–5 tell about mass changes in the Earth deep interior, while harmonics from 5–10 to 10–25 provide information on changes in the lower and upper mantle, respectively (Liu 1977; Bjerhammar et al. 1980). The lower and upper mantle contain about 50 and 10% of the Earth's mass, therefore studying those mass changes and effects on the geoid are important. The lower mantle extends from 650 to 2900 km below Earth's surface and the upper mantle from the Moho boundary to a depth of about 400 km (Table 8.1). Higher degree harmonics mainly tell about mass changes in the crust.

Figures 8.43 and 8.44 show the geoid (using 2.6f) and mass changes ($\dot{\kappa}$) in Fennoscandia, Antarctica and North America, where the spherical harmonic windows 5–10 and 10–25 have been applied, respectively. The contributions of the lower and upper mantle mass anomalies are different, mostly because of the ice-cap size, as the GIA signals in North America and Antarctica are much larger than in Fennoscandia. Generally, the same pattern can be seen for the geoid and mass changes in both lower and upper mantle, namely, when the geoid goes up, the mass change increases. The mass and geoid changes can be clearly observed in West Antarctica and Greenland, implied by on-going ice-mass loss. The signature of land uplift in the lower and upper mantle can be seen in Fennoscandia (Baltic Sea) and North America (Hudson Bay). The ice sheet in North America was much bigger in size, especially in its horizontal extent, than the one in Fennoscandia. It was, in some parts, also thicker than in Fennoscandia, although the difference in thickness would not explain all of the difference in the signal that can be seen today. Nevertheless, the whole load of the North-American ice cap (due its extent) was much bigger than the Fennoscandian, thus the lithosphere was pressed further down into the mantle. Also

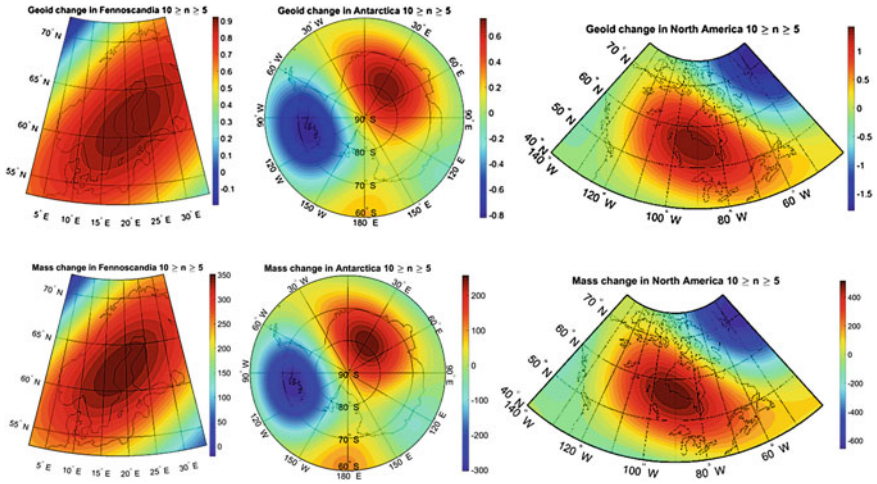

Fig. 8.43 Geoid (unit: mm/year) and mass changes (unit: kg/m^2/year) in Fennoscandia, Antarctica and North America for the spectral window $5 \leq n \leq 10$ (showing mostly changes in the lower mantle) using GRACE data for the time period of data between April 2002 and October 2014 (about 12.5 years)

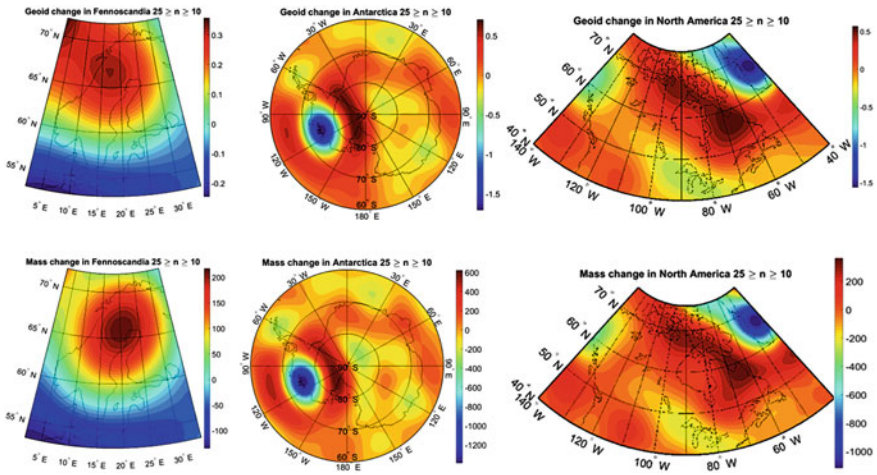

Fig. 8.44 Geoid (unit: mm/year) and mass changes (unit: kg/m^2/year) in Fennoscandia, Antarctica and North America for the spectral window $10 \leq n \leq 25$ (showing mostly changes in the upper mantle) using GRACE data for the time period of data between April 2002 and October 2014 (about 12.5 years)

note that the last remnants of the ice vanished later in North America than in Fennoscandia. Therefore, the mass changes and rebound were at a slightly different time stage in North America than in Fennoscandia (Steffen et al. 2010).

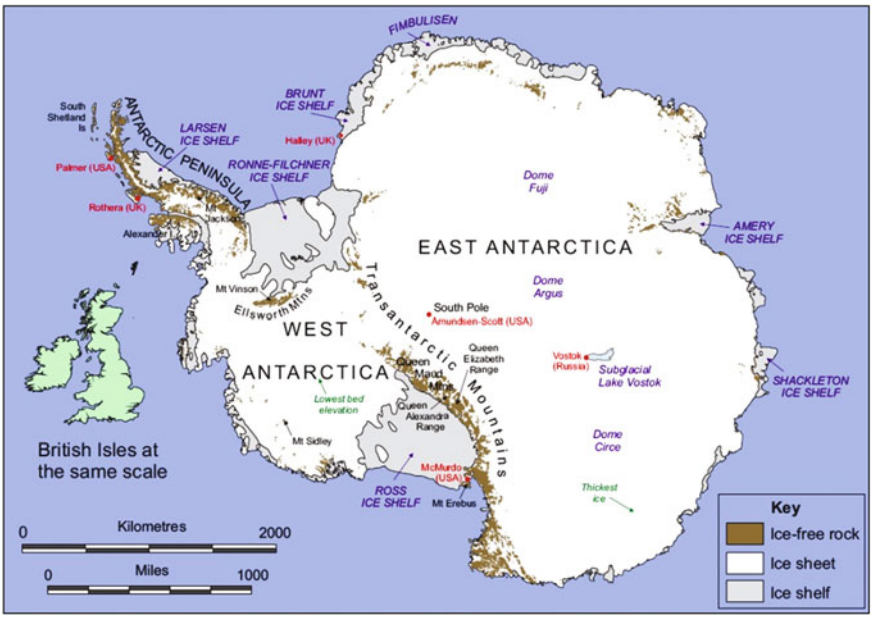

Fig. 8.45 Map of Antarctica (British Antarctic Survey)

The ice masses have a significant contribution to the gravity field in Antarctica, where it is estimated that the land is pushed down about 500 m by the ice. Figure 8.45 shows the names of major parts of Antarctica. The area is around 58 times the size of the UK, or 1.4 times the size of the USA. About 98% of Antarctica is covered by ice, which averages at least 2 km in thickness, and ice-free areas are only 0.32%. The maximum thickness of the ice is 4776 m in position 69°56'S, 135° 12'E. The three largest mountain ranges on the Antarctic continent are the Transantarctic Mountains, the West Antarctica Ranges and the East Antarctica Ranges. The Transantarctic Mountains compose a mountain range in Antarctica, which extends, with some interruptions, across the continent from Cape Adare in northern Victoria Land to Coats Land. These mountains serve as the divider between East and West Antarctica. They include a number of separately named mountain groups, which are often subdivided into smaller ranges. The geoid change exhibits a strong negative value of −4.5 mm/year, using about 12.5 years data, along the coast of West Antarctica (see Fig. 8.41a). This is a result of the fact that the glaciers in this region are losing significant mass (Sasgen et al. 2005), and the mass change is larger than in Fennoscandia and North America. However, the geoid changes in the lower and upper mantle are different than −4.5 mm/year (Figs. 8.43 and 8.44). The figures show that the geoid changes reach to −1.7 and −0.85 mm/year in the upper and lower mantle in West Antarctica (in 12.5 years), respectively. The findings from GOCE and GRACE and the data from ESA's

CryoSat satellite, which carries a radar altimeter, and also ICESat missions show a large negative ice-mass change in West Antarctica, a loss that has increased threefold since 2009.

In this chapter, the harmonic window technique was used to show the crustal, upper and lower mantle structures and mass changes beneath the above mentioned regions. Convective materials circulate in the mantle because of the temperature difference between the Earth's surface and outer core. Hot material rises from below, while cooler (and heavier) material descends. The latter is often in the form of large-scale lithospheric material moving downward at plate boundaries (subduction zones at about 410 km of depth). In the sense of fluid dynamics, mantle convection is a disordered process, which is supposed to be an integral part of plate motion.

8.8 Viscosity in the Mantle

The viscosity of a material is a measure of its resistance to gradual deformation by stress caused by an external force, mass change or load. Knowledge of the mantle viscosity is of vital importance in seismology, geology and studies on the GIA. Studies about GIA and viscosity determination are crucial, because the GIA causes changes in the shape, gravity field of the Earth and the Earth's rotation, as well as in sea level change. The mantle is a dynamo with flow driven by both thermal and chemical heterogeneities within it (Kaufmann and Lambeck 2000), and usually the travel times of seismic waves are used to infer its elastic properties such as viscosity. One way of estimating the mantle viscosity is to study the uplift process related to post-glacial rebound by geodetic and geologic information.

8.8.1 Geophysical Approaches

GIA modelling using a geophysical approach depends on reconstructions and assumptions of the glacial history and on the parameterization of the rheology of the solid Earth, mostly with rather limited data. For this purpose, various models of ice history and viscosity, e.g. ICE-5G (VM2) (Peltier 2004), the effect of Earth rotation (Milne and Mitrovica, 1998), an elastic lithosphere thickness model, are applied. There are three pillars in such a GIA modelling: observations, the ice model and the Earth model. Geologic, geophysical and/or geodetic data and topography (model) are the most important observations. Ice models show spatial variation over time that can be obtained based on geological evidence. The main problem is that the needed thickness of the ice in the models is difficult to determine accurately. The ice thickness depends on an Earth model that represents the structure of the Earth's crust and mantle. Different techniques are available for this purpose [see, for example, Wu et al. (2010), Spada et al. (2011), Steffen et al. (2012), and Peltier (2014)],

and all need an ice model and data for determining a realistic Earth model. Some benchmark projects for GIA modelling have been performed by Zhong et al. (2008) and Spada et al. (2011) to test different ice model effects on the GIA modelling techniques.

8.8.2 Rheology and Its Relationship to Viscosity

GIA modelling using geophysical approaches needs to assume a model for the Earth's rheology. Rheology is mainly related to continuum mechanics to characterize the flow of materials that exhibits a combination of elastic, viscous and plastic behaviour by appropriately combining elasticity and (Newtonian) fluid mechanics. It is evident that any displacement and geoid height changes depend on the way the Earth responds (e.g. by a rheology with either linear or non-linear models) to the surface load. Much of theoretical rheology is concerned with associating external forces and torques with internal stresses and internal strain gradients and flow velocities (Schowalter 1978). Assuming different rheologies for Earth gives us different results. Next we list various assumptions for the rheology of the Earth. By assuming a constant stress:

- If the material strain increases linearly with increasing applied stress, then the material is purely **elastic**.
- If the deformation rate increases linearly with increasing applied stress, then the material is purely **viscous**.
- If the material is both elastic and viscous, then a **viscoelastic** deformation follows.

The relationship between stress and strain is expressed by the equations in Table 8.10 as the responses of the Earth to loads for various rheology models.

Finally, which rheology fits best for the Earth's mantle can be identified. Practically, there is a combination of elasticity and viscosity models called the Maxwell model that fits well for the Earth's mantle: for a short period of time (e.g. for studying seismic waves), an elastic mechanism; while, for a long time span, e.g. in studying mantle convection, it is rather a viscous mechanism. Simply considering a Newton body for the Earth and also assuming that the deformation is directly related to viscosity (King 1995), the viscosity can be defined as:

$$\eta = \sigma/\dot{\varepsilon}, \tag{8.152}$$

where σ is the stress and $\dot{\varepsilon}$ is the strain rate. The mantle viscosity can be determined by the shape of the uplift curve, and it allows one to obtain the decay time (or relaxation time, i.e. the time it takes for the uplift to adjust to $1/e$ of the original depression). This gives the order of magnitude of (for a Newtonian body) the mantle viscosity (Haskell 1935):

Table 8.10 Various Earth models (rheology)

Model	Fundamental law	Constitutive equations	Remarks
Elasticity (Hooke body)	Stress is proportional to strain	$\frac{\sigma}{\varepsilon} = constant \rightarrow \sigma = \mu\varepsilon$	μ is a material constant known as rigidity, or shear modulus (having the dimension of stress)
Viscosity (Newton body)	Stress is proportional to strain rate	$\frac{\sigma}{\dot{\varepsilon}} = constant \rightarrow \sigma = \eta\dot{\varepsilon}$	η is a material constant (dependents on *pressure* and temperature), known as viscosity
Linear viscoelasticity (Maxwell rheology)	A Maxwell material is a viscoelastic material having the properties both of elasticity and viscosity	Maxwell rheological equation: $\dot{\varepsilon} = \frac{\dot{\sigma}}{\mu} + \frac{\sigma}{\eta}$	

$$\eta \approx 10^{21}\text{Pa s}. \tag{8.153}$$

King (1995) presented different models for mantle viscosity using different constraints such as the geoid change, plate tectonics, etc.

8.8.2.1 Classical (Geophysical) Approach to Determine Mantle Viscosity: A Geophysical Approach

According to Heiskanen and Vening Meinesz (1958, p. 358), mantle viscosity can be determined by the following approach by assuming that the crust sinks elastically under the ice load. After the ice is melted, the elastic stress in the mantle disappears by the slow relaxation and viscous property of the mantle. Then, the following differential equation expresses the balancing of the elastic and buoyancy forces versus the load (Lowrie 2007):

$$F_r\left\{\frac{\partial^4\xi}{\partial x^4} + 2\frac{\partial^4\xi}{\partial x^2\partial y^2} + \frac{\partial^4\xi}{\partial y^4}\right\} + (\rho_m - \rho)g\xi = L(x, y), \tag{8.154a}$$

where

$$F_r = \frac{E}{12(1-\nu^2)}D^3. \tag{8.154b}$$

Here, F_r is flexural rigidity, E is Young's modulus, ν is Poisson's ratio, D is the crustal plate thickness, x and y are horizontal plane coordinates, L is surface load's

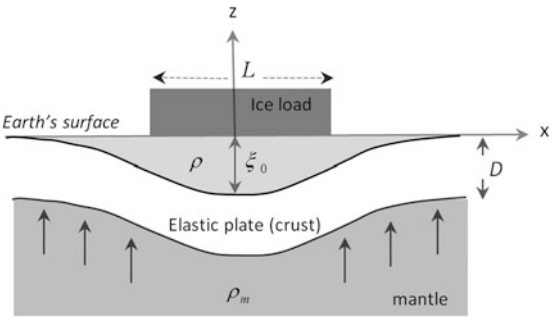

Fig. 8.46 Geometry of elastic bending of the crust due to ice load. D is the thickness of the plate (crust)

width, g is gravity, ρ is the density of the material that fills in the depression caused by the crustal deflection ξ, and ρ_m is density of the viscous mantle. If we assume that the load affects the plate only in x direction, the above differential equation becomes:

$$F_r \frac{\partial^4 \xi}{\partial x^4} + (\rho_m - \rho_c) g \xi = L(x). \tag{8.155}$$

Assuming that the crustal deflection vanishes in the x direction, the solution becomes (Heiskanen and Veining Meinesz 1958, p. 358):

$$\xi = \xi_0 \sin\left(\pi \frac{x}{L}\right), \tag{8.156}$$

which is a sinusoidal function, where ξ_0 is the maximum deflection along the ice axis. Figure 8.46 shows the geometry of bending of the elastic crust due to the ice load. The surface load L causes a deflection ξ_0 in the centre of the load.

When the ice disappears, the upward pressure becomes:

$$p = \rho_m g \xi. \tag{8.157}$$

The moment (M) can be obtained from the magnitude of the force multiplied by the distance. The force due to the load would cause the moment:

$$M = \rho_m g \frac{L^2}{\pi^2} \xi, \tag{8.158}$$

in the crust.

8.8.2.2 Mantle Viscosity

Taking advantage of the isostatic adjustment and assuming a viscous flow in the mantle, the relation between pressure due to load and the velocity components

(u, v, w) in the (x, y, z) directions due to the crust's load is given by the following Navier-Stokes equations (Heiskanen and Vening Meinesz 1958. pp. 358–361):

$$\begin{aligned}\rho_m \frac{\partial u}{\partial t} &= -\frac{\partial p}{\partial x} + \eta \Delta^2 u \\ \rho_m \frac{\partial v}{\partial t} &= -\frac{\partial p}{\partial y} + \eta \Delta^2 v \\ \rho_m \frac{\partial w}{\partial t} &= -\frac{\partial p}{\partial z} + \eta \Delta^2 w + \rho g,\end{aligned} \tag{8.159}$$

where Δ is the Laplace operator.

In this equation, it has been assumed that the mantle is plastic with limited elasticity, and the velocities are small (in the order of 1 cm/year). Therefore, higher-order terms can be neglected. Assuming that mass conservation holds:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \tag{8.160}$$

and using Eq. (8.159), the relation between stress and velocity components will be given by:

$$\begin{aligned}\sigma_x &= -p + 2\eta \frac{\partial u}{\partial x} \\ \sigma_y &= -p + 2\eta \frac{\partial v}{\partial y} \quad \textit{normal stress} \\ \sigma_z &= -p + 2\eta \frac{\partial w}{\partial z}\end{aligned} \tag{8.161a}$$

and

$$\begin{aligned}\tau_x &= \eta \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right) \\ \tau_y &= \eta \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right) \quad \textit{shear stress} \\ \tau_z &= \eta \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)\end{aligned} \tag{8.161b}$$

where the normal stress (compression or tension) is perpendicular to the surface, and the shear stress is parallel to the surface. Here, it has been assumed that the distribution of the mass change (loading or unloading) in the crust has a harmonic characteristic in x and y direction, and it has also been assumed that the mass change will slowly approach equilibrium. Therefore it can be assumed that u, v and w are functions of z multiplied by the following equation:

$$K\mathrm{e}^{-\upsilon t} = \cos(lx)\cos(my)\mathrm{e}^{-\upsilon t}, \tag{8.162}$$

where υ is decay parameter, which is the inverse of the decay time (see Sect. 8.8.6) and K is a harmonic function of x and y, which satisfies the harmonic equation

$$\nabla^2 K + f^2 K = 0 \quad \text{with} \quad f^2 = l^2 + m^2. \tag{8.163}$$

Using these assumptions, the solution of velocity and stress components that satisfies Eq. (8.160) will be obtained as (Heiskanen and Vening Meinesz 1958, pp. 358–362):

$$\begin{aligned} u &= Az\frac{\partial K}{\partial x}\mathrm{e}^{-fz-\tau t} \\ v &= Az\frac{\partial K}{\partial y}\mathrm{e}^{-fz-\tau t} \\ w &= -A(fz+1)K\mathrm{e}^{-fz-\tau t} \end{aligned} \tag{8.164}$$

and

$$\begin{aligned} \sigma_x &= -\rho g z + 2\eta A\left(fK + z\frac{\partial^2 K}{\partial x^2}\right)\mathrm{e}^{-fz-\tau t} \\ \sigma_y &= -\rho g z + 2\eta A\left(fK + z\frac{\partial^2 K}{\partial y^2}\right)\mathrm{e}^{-fz-\tau t} \\ \sigma_z &= -\rho g z + 2\eta A(1+fz)fK\mathrm{e}^{-fz-\tau t}, \end{aligned} \tag{8.165}$$

where A is an integration constant that can be derived from Eq. (8.160).

The relation between the vertical depression of the crust and the viscosity can be obtained by the following assumption:

$$\lim_{z\to 0}\sigma_z = \xi, \tag{8.166}$$

Implying that for $z = 0$, σ_z equals the crustal deflection multiplied to ρg (deviation from equilibrium):

$$\xi = \frac{2\eta f}{\rho g}AK\mathrm{e}^{-\tau t}. \tag{8.167}$$

Also, the differential ξ of with respect to time is equal to w (i.e. the uplift rate) when $z = 0$:

$$-\frac{2\eta f}{\rho g}\tau AK\mathrm{e}^{-\tau t} = -AK\mathrm{e}^{-\tau t}, \tag{8.168}$$

therefore, the viscosity follows by:

$$\eta = \frac{\rho g}{2\tau f}, \tag{8.169a}$$

and the decay parameter:

$$\tau = \frac{\dot{\xi}}{\xi}\frac{\rho g}{2\eta f}. \tag{8.169b}$$

Here, the decay parameter τ is obtained by dividing the land-uplift rate by the postglacial depression. Figure 8.47 shows the relationship between the depression and the parameter τ in Fennoscandia. The earliest value of depression (556 m) is derived by Nansen (1928) for 8000 years B.C., and we can assume 137 m in year 2000 on the west coast of the Gulf of Bothnia.

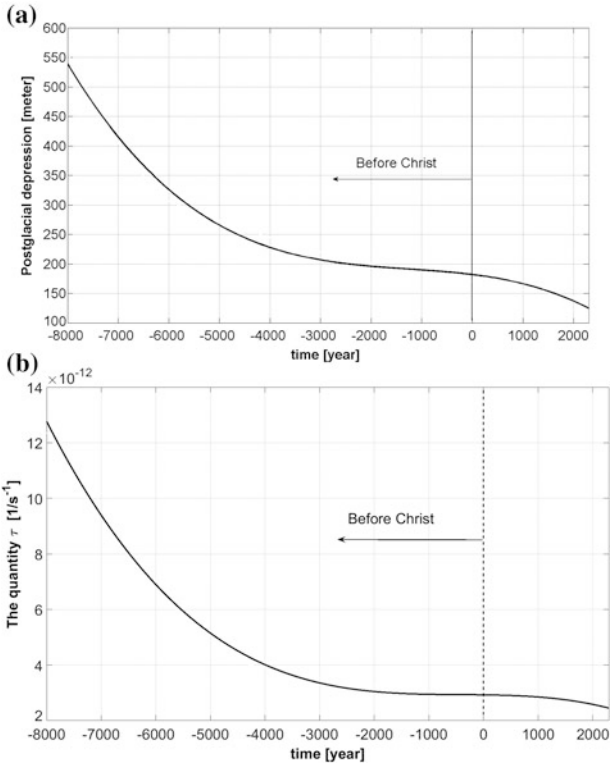

Fig. 8.47 Postglacial depression and decay parameter τ in Fennoscandia on the west coast of the Gulf of Bothnia (Heiskanen and Vening Meinesz 1958)

Assuming $f = \pi/L = \pi/(1.4 \times 10^8)$, which corresponds to the ice dimension 1400 km in plane direction and 2300 km in vertical direction, ρ_m = 3.27 g/cm^3 and $\tau = 6 \times 10^{-12}\text{s}^{-1}$, the viscosity will be obtained as:

$$\eta = 1.02 \times 10^{22}\text{Pa s} \tag{8.170}$$

More recent methods for determining the viscosity can be found in Spada et al. (2011) using Maxwell rheology and the load-deformation technique.

8.8.3 A Geodetic Approach

Following Sjöberg and Bagherbandi (2013) and assuming that the elastic deformation of the crust of thickness D and density ρ_c is negligible in the vertical direction, the contribution to the disturbing potential at sea level, due to the glacial depression of the crust, can be expressed as:

$$\Delta T = -G\rho_c \iint_{\sigma} \int_{R-d}^{R} \frac{r^2 dr}{l} d\sigma - G\Delta\rho \iint_{\sigma} \int_{R-D-d}^{R-D} \frac{r^2 dr}{l} d\sigma, \tag{8.171}$$

where G is the gravitational constant, $\Delta\rho$ is the difference between the mantle and crust densities, σ is the unit sphere, R is the sea level radius, l is the distance between the computational point and integration point and d is the depression of the crust, or, it can be regarded as the remaining uplift (if it will be completely restored to sea level). The first integral is the effect on the Earth's potential due to the mass deficit between sea level and the depressed crust. The second term is the gravitational effect due to the mass deficit caused by the depressed bottom of the crust into the mantle. We assume that the upper and lower surfaces of the crust are equally deformed and depressed by the glaciation. Expanding the inverse distance in an exterior type of Legendre series, one obtains (to first-order Taylor series approximations):

$$\begin{aligned} \Delta T &\approx -\mu R \sum_{n=0}^{\infty} \iint_{\sigma} dP_n(\cos\psi) d\sigma = -4\pi\mu R \sum_{n=0}^{\infty} \frac{d_n}{2n+1} \\ &= -4\pi\mu R \sum_{n=0}^{\infty} \frac{1}{2n+1} \sum_{m=-n}^{n} d_{nm} Y_{nm}, \end{aligned} \tag{8.172}$$

where $\mu = G\rho_m$, and in the last step we have utilized the addition theorem Eq. (2.49).

Note that ρ_m is the density of the upper mantle. From Eq. (8.172), we thus obtain the following spectral relation between the disturbing potential and the remaining uplift:

$$\Delta T_{nm} \approx -4\pi\mu R \frac{d_{nm}}{2n+1} \quad \text{or} \quad \Delta T_n \approx -4\pi\mu R \frac{d_n}{2n+1}. \tag{8.173}$$

8.8.4 *Estimating the Remaining Land Uplift from the Geoid Depression*

Applying Bruns' formula for the relationship between the fraction of the disturbing potential ΔT and the geoid height ΔN related with GIA:

$$\Delta N = \frac{\Delta T}{\gamma_0}, \tag{8.174}$$

where γ_0 is normal gravity on the reference ellipsoid, it follows from Eq. (8.173) that:

$$d \approx -c \sum_{n=n_{\min}}^{n_{\max}} (2n+1) N_n; \quad \text{with} \quad c = \frac{\gamma}{4\pi\mu R} \approx 0.5567. \tag{8.175}$$

Following Bjerhammar et al. (1980), we will use the spectral window $n_{\min} = 10$ and $n_{\max} = 23$, which is the spectral window with optimum correlation between the land uplift rate and gravity potential in the region (see Sect. 8.8.6 for further details).

The disturbing potential is computed as the difference between the Earth's gravitational potential and normal potential given by GOCO02S (Goiginger et al. 2011) and the GRS-80 reference field (Moritz 2000), respectively. The computations are realized on a 1 × 1 arc-degree geographical grid of surface points. The normal gravity at the reference ellipsoid (γ_0) is approximated to 981 Gal. Figure 8.48 shows the corresponding geoid height depression in Fennoscandia, which varies between −8.9 and 0.19 m, with a mean value of −5.5 ± 2.1 m.

The numerical result for the remaining uplift from the present time up to completeness is shown in Fig. 8.49. The maximum remaining uplift of 134 m is estimated on the Bothnian Bay and the mean value is 76 m in Fennoscandia. Some previous studies estimated the remaining uplift to 90 m (Sjöberg et al. 1994), 50–90 m (Ekman 1991), 90 m (Ekman and Mäkinen 1996), 55–90 m (Kaufmann et al. 2000), while Fjeldskaar and Cathles (1991) suggested a remaining maximum uplift of only 40 m.

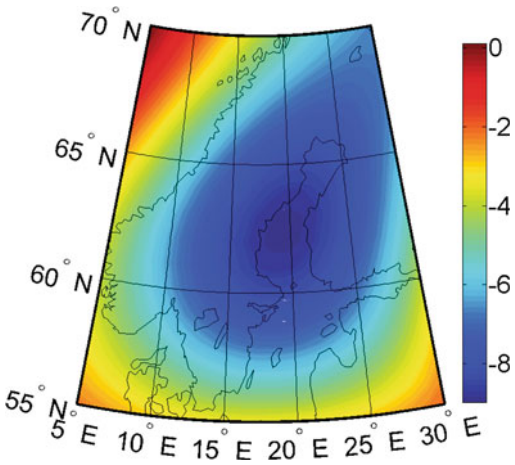

Fig. 8.48 The geoid over Fennoscandia for spectral window $10 \leq n \leq 23$. Unit: metre

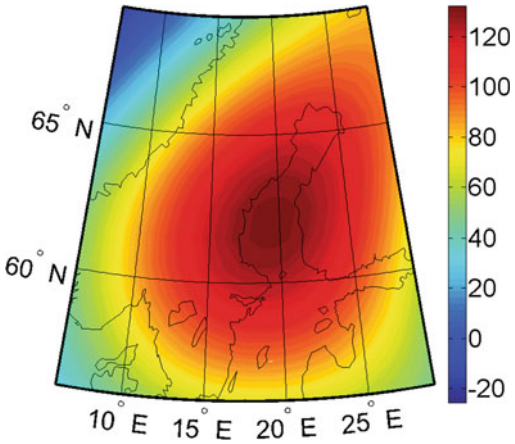

Fig. 8.49 Remaining land uplift for spectral window $10 \leq n \leq 23$ in Fennoscandia. Unit: metre

8.8.5 Estimating the Geoid Height and Absolute Uplift Rates

The absolute land uplift rate $\dot{h}$ (with respect to the gravity centre of the Earth) is the sum of the observed/apparent uplift rate (with respect to mean sea level; $\dot{H}$), the eustatic change of sea level (caused by melting of ice caps; changes of sea bottom topography and temperature related expansion of seawater; $\dot{H}_e$) and the rate of change of the geoid height ($\dot{N}$):

$$\dot{h} = \dot{H} + \dot{H}_e + \dot{N} \tag{8.176}$$

Numerically, we will use the apparent uplift model NKG2005LU, originally compiled by Ågren and Svensson (2007) and (2011) from geodetic observations (55 continuously operating GPS and 56 tide gauge stations and a few repeated national precise-levelling lines), and, in areas without geodetic data, the preferred GIA model (constrained by the above-tide gauge data) of Lambeck et al. (1998). According to J. Ågren (private communication), the impact of the GIA model is only minor, implying that the land-uplift model is mainly based on geodetic data. The eustatic uplift rate is about 1.32 mm/year for the considered period 1882–1992 (Ågren and Svensson 2011).

We now assume that the post-glacial rebound process is the reverse of the glacial depression process. Taking the time derivative of both sides of Eq. (8.172), noting that $\dot{h} = -\dot{d}$ and applying Bruns' formula, one obtains:

$$\dot{N} \approx q \sum_{n=0}^{\infty} \frac{\dot{h}_n}{2n+1}, \tag{8.177a}$$

where:

$$q = \frac{4\pi\mu R}{\gamma} = 1/c \approx 1.796. \tag{8.177b}$$

Considering also Eq. (8.176), and that the eustatic effect is only a zero-degree effect, one arrives at the following spectral relations for the geoid height and absolute uplift rates:

$$\dot{N}_n = \frac{q}{(2n+1)-q}\dot{H}_n, \quad n \geq 1 \tag{8.178}$$

and

$$\dot{h} = \dot{H}_e + \sum_{n=n_{\min}}^{n_{\max}} \frac{2n+1}{2n+1-q}\dot{H}_n. \tag{8.179}$$

The smoothed absolute land-uplift rate, as depicted in Fig. 8.50, varies from 1.4 to 10.9 mm/year with a mean value of 4 mm/year and a standard deviation of 2.7 mm/year.

The estimated uplift rate of the geoid, shown in Fig. 8.51, has a peak value of 0.70 mm/year, which agrees well with 0.7 mm/year estimated by Sjöberg (1983) who also used land-uplift data, and also with the geoid change estimated from GRACE data, shown in Fig. 8.41.

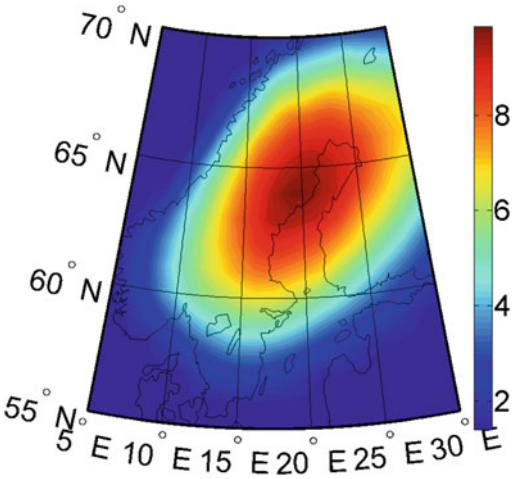

Fig. 8.50 Absolute land uplift rate model based on NKG2005LU (Ågren and Svensson 2007). Unit: mm/year

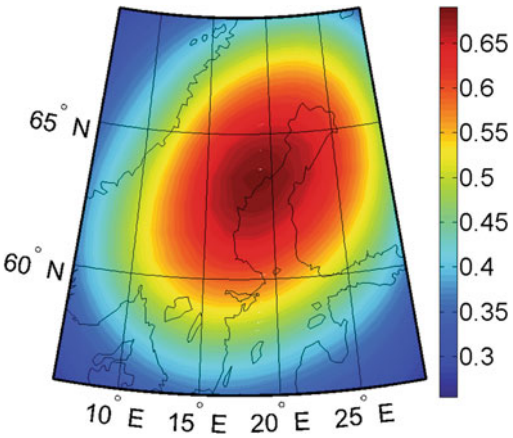

Fig. 8.51 Temporal change of the geoid estimated from the land-uplift rate in Fennoscandia. Unit: mm/year

8.8.6 The Decay Time

Assuming a spherical Earth with a solid crust floating on a viscous upper mantle, the remaining uplift at time t is related to the total depression h_0 at time $t = 0$ by the spectral relation (e.g. de Geer 1888; Niskanen 1949; Walcott 1980) so:

$$h_n = (h_0)_n e^{-t/\tau_n}, \tag{8.180}$$

where τ_n is the spectral decay or relaxation time (i.e. the inverse of the decay parameter), which is the time it takes for the uplift to adjust to $1/e$ of the original depression, where $e \approx 2.67$ is the base of the natural logarithm. By differentiating this equation with respect to time and introducing uplift rate $\dot{h}_n$ one easily obtains:

$$-\dot{d}_n = \dot{h}_n = d_n/\tau_n, \tag{8.181}$$

And, by inserting d_n from the spectrum of Eq. (8.175), the decay time can be expressed as:

$$\tau_n \approx -(2n+1)c\frac{N_n}{\dot{h}_n}. \tag{8.182}$$

Figure 8.52a, b show the geoid height and uplift rate, respectively, degree by degree. Comparing the figures, we observe that there is a negative correlation between the two plots above degree 10 (which we assume is the lower limit of the glacially isostatically induced geoid depression). The total correlation between the geoid height and land-uplift rate for the considered spectral window is -0.99 ± 0.006. The decay time as given by Eq. (8.182) and shown in Fig. 8.52c is generally increasing (almost linearly) by degree. (Also, notice the unexplained disturbances of the linear trend with local high and low peaks at degrees 45, 52 and 67 corresponding to wavelengths of about 220, 190 and 148 km, respectively, which are outside the optimum GIA related spectral window $10 \leq n \leq 23$). Comparing with previous results, based on shoreline data, we obtain considerably longer relaxation times. For example, Watts (2001, p. 115), based on a study by Sauramo (1958), presented an estimate of the decay time of 8 kyr for (the dominating) degree 16, while Fig. 8.52c suggests 6 kyr for this degree. [To understand this discrepancy, one should bear in mind that we determine the gravity spectrum from a precise Earth gravitational model based on satellite gravity, while Sauramo (1958) only had access to much poorer terrestrial gravity data.]

Traditionally there are two simple models for the GIA mantle flow in Fennoscandia. According to Model I (thin-channel flow) the decay time should increase by spatial wavelength, while the opposite holds for Model II (whole-mantle flow). This method to distinguish between the two general models was first studied by McConnell (1965) and (1968) by using Fennoscandian uplift data, but the data were not sufficiently precise to come to a firm conclusion. Other similar attempts have been conducted with Laurentide uplift data and also by comparing the integrated relaxation times vs. size of the areas of the Laurentide and Fennoscandian post-glacial rebounds, but again, the result were not conclusive, probably due to imprecise historical shoreline data. On the other hand, our computations based on recent uplift data from Fennoscandia, illustrated by Fig. 8.52c, clearly show that the decay time decreases (almost linearly) with wavenumber (n),

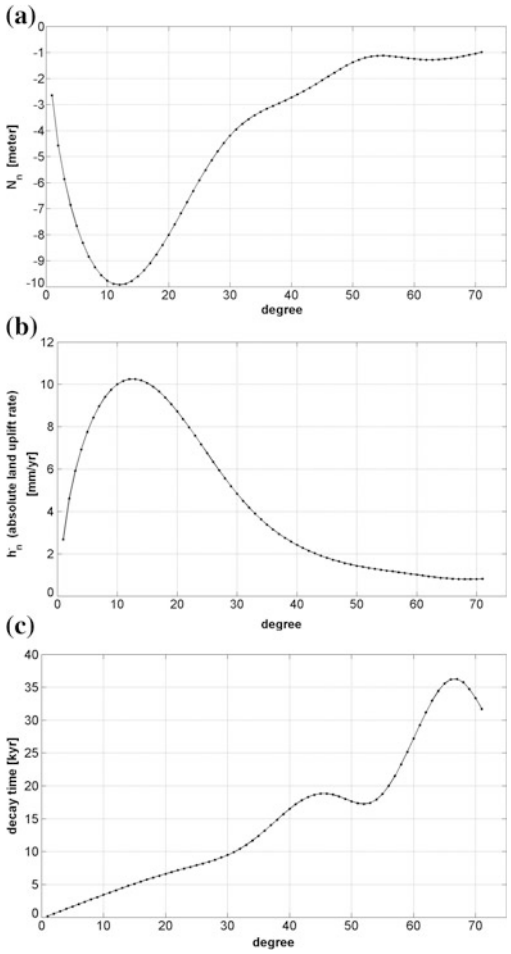

Fig. 8.52 **a** N_n, **b** $\dot{h}_n$ (absolute land-uplift rate) and **c** decay time

which supports Model II (see also Eq. (8.183) in the next subsection). This result also agrees with Fig. 8.53, which shows that the viscosity increases by wavelength (except for some minor disturbances).

8.8.7 Remaining Uplift Versus Power of Uplift Rate

According to Walcott (1980), the shapes of the uplift curves are significantly different for Models I and II due to the differences in viscous mantle flow. This difference can be explained because the uplift rate is linearly and cubically

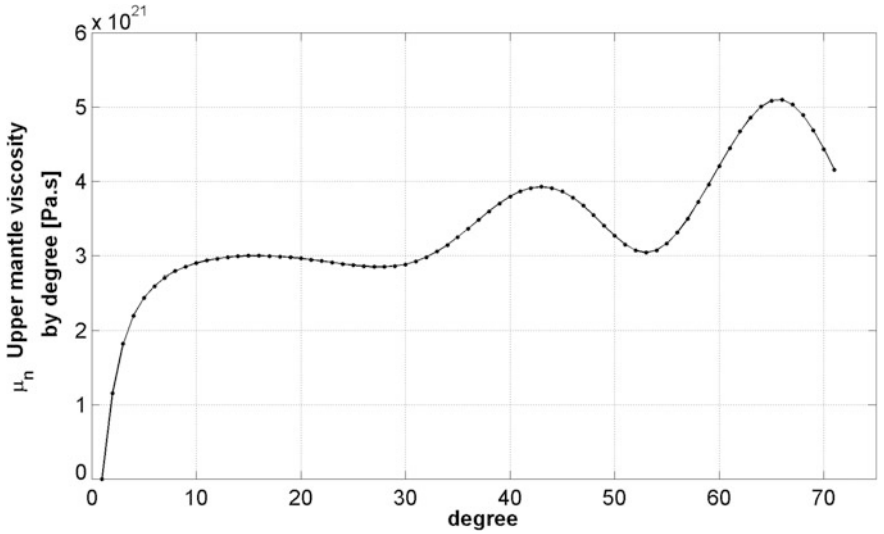

Fig. 8.53 Upper-mantle viscosity by degree. Unit: Pa s

proportional to the remaining uplift, for Models II and I, respectively. However, as the remaining uplift is not an observable quantity, these patterns are difficult to use in making distinctions between the two models. However, the remaining uplift is closely related with the glacio-induced geoid depression (see Eq. 8.175), which is an observable. Hence, based on our previous computational results, linear and cubic correlations of uplift rate with remaining uplift and also with the geoid depression were computed. The computational results yielded correlation coefficients of 0.63 and 0.44 for the linear (related with Model II) and cubic (related with Model I) correlations, respectively (with no significant difference for remaining uplift or geoid depression as input variables). Hence, once again, the computational results support Model II.

8.8.8 Upper-Mantle Viscosity

For the deformation of a viscous sphere, the decay time τ_n and viscosity η are related by the expression (e.g. Cathles 1971, 1975; Peltier 1974):

$$\tau_n = \frac{\eta_n}{\gamma \rho_m R}\left(2n + 4 + \frac{3}{n}\right), \tag{8.183}$$

and by inserting Eq. (8.182) one arrives at the following estimate for the viscosity (Sjöberg and Bagherbandi 2013):

$$\eta_n \approx -\frac{\gamma^2}{4\pi G}\frac{2n+1}{2n+4+3/n}\frac{N_n}{\dot{h}_n}. \tag{8.184a}$$

As the postglacial rebound in Fennoscandia is a regional phenomenon mainly related to harmonics of degrees, say, larger than 10 (Bjerhammar et al. 1980), the ratio $(2n+1)/(2n+4+3/n)$ in the equation is nearly one for all degrees $n \geq 10$, which yields:

$$\eta_n \approx -\frac{\gamma^2}{4\pi G}\frac{N_n}{\dot{h}_n}. \tag{8.184b}$$

In Fig. 8.53, we present the regional averages values for the upper mantle viscosity by degree using Eq. (8.184b), which shows an almost constant viscosity by degree [in the range of (4–7) $\times$ 10^{21} Pa s] for the studied spectral window.

If we further assume that the ratio $N_n/\dot{h}_n$ is independent of the wavelength, it follows that:

$$\eta \approx -\frac{\gamma^2}{4\pi G}\frac{N}{\dot{h}}. \tag{8.184c}$$

Hence, viscosity of the upper mantle utilizing the absolute land uplift rate and a geoid model can be determined using Eq. (8.184c). (Note that the absolute land uplift rate must be converted from mm/year to mm/s by the factor 12 months $\times$ 30 days $\times$ 24 h $\times$ 3600 s.) The estimated viscosity derived from the land uplift rate varies from 1.1 $\times$ 10^{21} to 1.6 $\times$ 10^{22} Pa s over Fennoscandia, with a mean value in the central uplift region (within a radius of 3 arc-degree from the centre) of 4.3 $\times$ 10^{21} Pa s. The most rapid lateral change of the viscosity occurs at shallow depths of the crust, increasing from 10^{21} to 10^{22} Pa s in the Norwegian Sea and Finland. For comparison, as shown in Table 8.11, most recent studies estimate the upper-mantle viscosity in the range of (0.5–5) $\times$ 10^{21} Pa s, and our results are also in that range. However, in contrast to many of the methods in the comparison, from our limited areal extension of data and approximate method, we cannot expect to also obtain an estimate of viscosity at depth.

We can summarize this numerical experiment as follows:

Repeated absolute gravity measurements in Fennoscandia have revealed that the on-going post-glacial rebound can be regarded as a pure viscous flow of mantle mass of density 3390 kg/m^3 towards the central part of the region caused by a gravity/uplift rate of -0.167 μGal/mm. Our model estimates the rebound induced rates of changes of surface gravity and geoid height to have peaks of −1.9 μGal/year and 1.6 mm/year, respectively. The rates of change of gravity agree well with recent absolute gravity data. The results, based on a recent land uplift rate model in Fennoscandia and an up-to-date Earth gravitational model, confirms the

Table 8.11 Comparison of some estimates of [upper] mantle viscosity in Fennoscandia in units of 10^{21} Pa s

Reference	Model	Data	Viscosity
Niskanen (1949)	Elastic crust and viscous flow in the mantle	Historical shorelines	5
Bjerhammar et al. (1980)	Viscous flow in the mantle	EGM and $\dot{h}$	40
Fjeldskaar and Cathles (1991)	Viscoelastic asthenosphere and mantle	$\dot{h}$	1.2
Sjöberg et al. (1991)	Simple viscous flow	EGM, $\dot{h}$ and T	2
Sjöberg et al. (1994)	Regression analysis	EGM, $\dot{h}$ and T	2.2 ± 0.2
Kaufmann and Lambeck (2000)	Forward and inverse modelling	EGM, sea-level change, seismic tomography	0.2–0.5 upper 10–30 lower
Milne et al. (2001)	Glacial-isostatic adjust in Fennoscandia	GPS and tide gauge data	0.5–1 upper 10 lower
Bergstrand et al. (2005)	Ice model RSES	GPS	30
Steffen et al. (2010)	Ice model RSES	GRACE	0.2–0.4
Lidberg et al. (2010)	Ice model RSES	GPS	0.5 upper 5 lower
Zhao et al. (2012)	Inverse modelling	Ice sheet models and $\dot{h}$ from GPS	0.3–0.5 upper 7–13 lower
Sjöberg and Bagherbandi (2013)	Simple viscous flow	EGM, $\dot{h}$	4 for spectral window $10 \leq n \leq 70$
This study	Simple viscous flow	EGM, $\dot{h}$	4 for spectral window $10 \leq n \leq 23$
This study	Simple viscous flow	EGM, $\dot{N}$	6.5 for spectral window $10 \leq n \leq 23$

T = Moho depth data. $\dot{h}$ = present day uplift rate data (repeated levelling, GPS and tide gauge data)

viscous flow model in the mantle (in contrast to the thin-channel flow model in the upper mantle). The result is supported both by the decreasing relaxation time with spatial wavelength of uplift rate and by the type of correlation between remaining uplift (or present geoid depression) and uplift rate. The upper mantle viscosity is estimated to 6×10^{21} Pa s, with a small increase in the range of $(1–16) \times 10^{21}$ Pa s by degree in the spectral window studied. One may speculate that the physics

behind the decreasing relaxation time by wavelength, as well as the shape of the viscosity curve in Fig. 8.53, support rheology models with increasing mantle viscosity with depth, but, considering the limited area of data and the approximate methodology, such a conclusion is out of the scope in this study. Here, GRACE data could help (see the next section).

8.8.9 Viscosity Determination Using GRACE Data

The most successful data for the temporal changes of the gravity field stems from the accurate dedicated satellite gravity mission GRACE. The GRACE monthly solutions have been used in many studies to infer mantle rheology and ice history in areas with GIA phenomena (e.g. Steffen et al. 2010).

In this section, viscosity determination is presented using the temporal change of the geoid change obtained from GRACE data. The inverse of Eq. (8.177a) becomes:

$$\dot{h} \approx c \sum_{n=0}^{\infty} (2n+1)\dot{N}_n; \quad c \approx 0.5567, \tag{8.185}$$

which, inserted into Eq. (8.184c), yields:

$$\eta \approx -\frac{\gamma^2}{4\pi G c} \frac{N}{\sum_{n=0}^{\infty}(2n+1)\dot{N}_n} = -\gamma \rho_m R \frac{N}{\sum_{n=0}^{\infty}(2n+1)\dot{N}_n}; \quad \mu = G\rho_m \quad \text{and} \quad \frac{4\pi\mu R}{\gamma} = 1/c. \tag{8.186}$$

We infer the viscosity of the Earth's mantle from observations of long-wavelength geoid N, the harmonic window $n_{\min} = 10$ and $n_{\max} = 23$ (see Sect. 8.8.3) and employ secular geoid changes ($\dot{N}_n$) obtained from GRACE data (see Sect. 8.7). Using Eq. (8.186), the viscosity is obtained to 6.5×10^{21} Pa s in the central uplift region (within a radius of 3° of the centre). This result is comparable with the results presented in Sect. 8.8.8. This result, based on GRACE data, is related to the upper-mantle layers down to the transition zone at a 660 km depth according to the schematic viscosity profile presented by Kaufmann and Lambeck (2000), who reported average viscosities of about (2–5) $\times 10^{20}$ Pa s in the upper mantle, and (1–3) $\times 10^{22}$ Pa s in the lower mantle, which are similar to earlier inferences by Forte and Mitrovica (1996). Previous studies show an almost uniform viscosity profile (e.g. Cathles 1975; Peltier and Andrews 1976), while most studies also suggested a significant one-to–two order- magnitude increase of viscosity across the 660 km seismic discontinuity (Forte and Mitrovica 1996; Lambeck et al. 1996, 1998). As can be observed by the profiles in Fig. 8.54, the viscosity

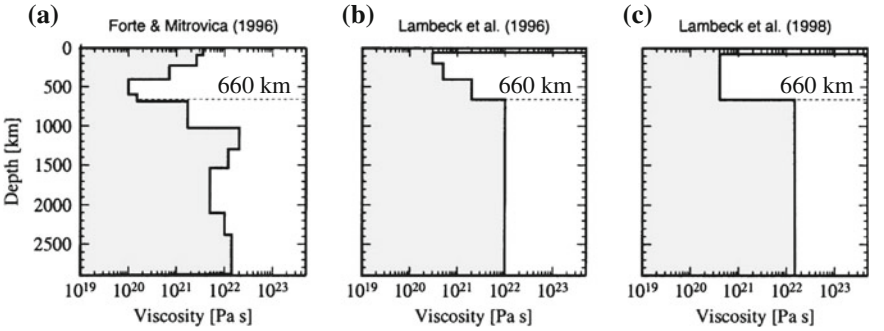

Fig. 8.54 Viscosity profiles from predictions of postglacial signatures (Kaufmann and Lambeck 2000)

variations can be smooth in the mantle (profile b and c) or a very large viscosity jump across the 660 km discontinuity (profile a).

The above mentioned studies modelled the time dependent glacial isostatic adjustment of the Earth related to past and present changes in the ice ocean mass imbalance based on mantle viscosity profiles. Also they deduced mantle-viscosity profiles, which simultaneously fit the long-wavelength geoid constraint and glacially induced changes of the Earth's shape.

Finally, in Fig. 8.55, we show the land uplift rate in Fennoscandia estimated by GRACE data and Eq. (8.185). Comparing with Fig. (8.50), the maximum change is about 35% larger and located further to the NE.

As previously discussed, the Earth's gravity field changes with time are due to various geodynamical processes. The motions can be classified as secular and periodic, and our focus has been on studying secular trends. The largest secular

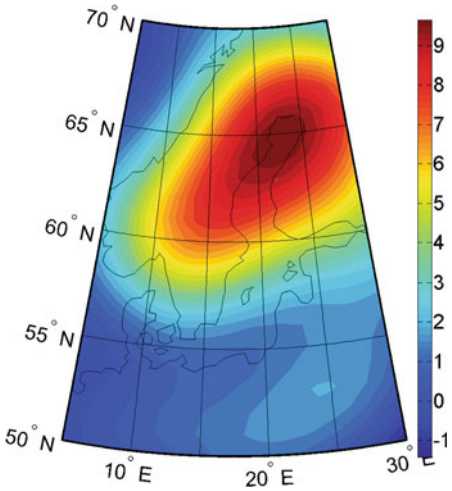

Fig. 8.55 Land uplift rate ($\dot{h}$) obtained from GRACE data using the CSR processing data centre (The University of Texas at Austin, Centre for Space Research) for the time period between April 2002 and October 2014 (about 12.5 years) after applying DDK4 filter (Kusche et al. 2009). Unit: mm/year

changes are caused by mantle convection, and plate and intraplate tectonics, as well as GIA. Today, the most successful data for studying the temporal changes of the gravity field stem from the long record of the accurate dedicated GRACE satellite gravity mission. These data greatly complement traditional GIA studies by geologic records and geodetic repeated levellings and GNSS records because of the long-wavelength characteristic of the phenomenon. One result of this section shows that some of the periodic terms are not significant in the estimation of the trend, and they can therefore be ignored (see Table 8.9). The obtained land uplift rates in Fennoscandia, based on the inversion of gravity data, are also comparable with the uplift rates calculated with GPS station records (Lidberg et al. 2010; Kierulf et al. 2014).

References

Abramowitz M, Stegun I (1964) Handbook of Mathematical functions. National Bureau of Standards, Washington, DC

Airy GB (1855) On the computations of the effect of the attraction of the mountain masses as disturbing the apparent astronomical latitude of stations in geodetic surveys. Trans Roy Soc (London), Ser B 145

Afonso JC, Fernàndez M, Ranalli G, Griffin WL, Connolly JAD (2008) Integrated geophysical-petrological modeling of the lithosphere and sublithospheric upper mantle: methodology and applications. Geochem Geophys Geosyst 9:Q05008

Ågren J, Svensson R (2007) Postglacial land uplift model and system definition for the new Swedish height system RH 2000. Reports in Geodesy and Geographical Information Systems, LMV-Rapport 2007:4. Lantmateriet, Gävle, Sweden

Ågren J, Svensson R (2011) The height system RH 2000 and the land uplift model NKG2005LU. Mapp Image Sci 2011(3):4–12

Allenby RJ, Sehnetzler CC (1983) United States crustal thickness. Tectonophysics 93(1983):13

Anderson DL (1989) Theory of the Earth. Blackwell Scientific Publications, Boston. http://resolver.caltech.edu/CaltechBOOK:1989.001

Artemieva IM, Mooney WD (2001) Thermal thickness and evolution of Precambrian lithosphere: a global study. J Geophys Res 106:16387–16414

Artemieva IM (2006) Global 1° × 1° thermal model TC1 for the continental lithosphere: implications for lithosphere secular evolution. Tectonophysics 416:245–277

Artemieva IM, Thybo H, Kaban M (2006) Deep Europe today: Geophysical synthesis of the upper mantle structure and lithospheric processes over 3.5 By, in European Lithosphere Dynamics, Gee DG, Stephenson RA (eds). Geol Soc Mem, Lond 32:11–41

Assumpção M, Bianchi MB, Julià J, Dias FL, França GS, Nascimento RM, Drouet S, Pavão CG, Albuquerque DF, Lopes AEV (2013) Crustal thickness map of Brazil: data compilation and main features. J S Am Earth Sci. doi:10.1016/j.jsames.2012.12.009. Accepted 18 Dec 2012

Bagherbandi M (2011) An isostatic Earth crustal thickness model and its applications. PhD thesis. Royal institute of technology, 207 p

Bagherbandi M, Sjöberg LE (2013) Improving gravimetric-isostatic models of crustal depth by correcting for non-isostatic effects and using CRUST2.0. Earth Sci Rev 29–39. doi:10.1016/j.earscirev.2012.12.002

Bagherbandi M, Tenzer R, Sjöberg LE, Novak P (2013) Improved global crustal thickness modeling based on the VMM isostatic model and non-isostatic gravity correction. J Geodyn 66 (2013):25–37. doi:10.1016/j.jog.2013.01.002

Bai Y, Williams S, Müller R, Liu Z, Hosseinpour M (2014) Mapping crustal thickness using marine gravity data: methods and uncertainties. Geophysics 79:G27–G36
Baranov A, Morelli A (2013) The Moho depth map of the Antarctica region. Tectonophysics 609:299–313. doi:10.1016/j.tecto.2012.12.023. 8 Dec 2013
Barzaghi R, Biagi L (2014) The collocation approach to Moho estimate. Ann Geophys 57(1): S0190. doi:10.4401/ag-6367
Bassin C, Laske G, Masters TG (2000) The current limits of resolution for surface wave tomography in North America. EOS Trans AGU 81:F897
Becker TW, O'Connell RJ (2001) Lithospheric stresses caused by mantle convection: the role of plate rheology. Am Geophys Union, Fall Meeting 2001, abstract #T12C-0921
Belyaevsky NA, Volkovsky IS (1980) Tectonic map of Northern Eurasia. In: Relief of the Moho Surface, Moscow
Belyaevsky NA (1981) Structure of the earth crust from geologic-geophysical data. Nedra, Moscow (in Russian)
Bendick R, Bilham R, Freymueller J, Larson K, Yin G (2000) Geodetic evidence for a low slip rate in the Altyn Tagh fault system. Nature 404(6773):69–72. doi:10.1038/35003555. PMID 10716442
Bentley CR, Wahr JM (1998) Satellite gravity and the mass balance of the Antarctic ice sheet. J Glaciol 44:207–213
Bercovici D (2010) Mantle convection. In: Gupta HK (ed) Encyclopedia of Solid Earth Geophysics. Springer 2011. 1539 Seiten
Bergstrand S, Scherneck H-G, Milne GA, Johansson JM (2005) Upper mantle viscosity from continuous GPS baselines in Fennoscandia. J Geodyn 39:91–109. doi:10.1016/j.jog.2004.08.004
Bjerhammar A, Stocki S, Svensson L (1980) A geodetic determination of viscosity. Division of Geodesy. The Royal Institute of Technology, Stockholm, Sweden
Bois P (1961) Indefinite integrals. Dover, New York
Bott MHP (1971) The interior of the Earth. Edward Arnold Ltd., London
Bouhifd MA, Andrault D, Fiquet G, Richet P (1996) Thermal expansion of forsterite up to the melting point. Geophys Res Lett 23:1143–1146. doi:10.1029/96GL01118
Bowin C (1994) The geoid and deep earth mass anomaly structure. In Vanicke P, Christou N T (Eds.): Gravity and its geophysical interpretation, CRC Press, Inc, Chap. 10
Bungum H, Pirhonen SE, Husebye ES (1980) Crustal thicknesses in Fennoscandia. Geophys J Int 63(3):759–774. doi:10.1111/j.1365-246X.1980.tb02650.x
Bürgmann R, Rosen PA, Fielding EJ (2000) Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. Annu Rev Earth Planet Sci 28:169–209
Čadak O, Martinec Z (1991) Spherical harmonic expansion of the earth's crustal thickness up to degree and order 30. Studia Geoph Et Geod 35:151–165
Cathles LM (1971) The viscosity of the Earth's mantle. PhD thesis, Princeton University
Cathles LM (1975) The viscosity of the Earth's mantle. Princeton University Press, Princeton
Chappell AR, Kusznir NJ (2008) Three-dimensional gravity inversion for Moho depth at rifted continental margins incorporating a lithosphere thermal gravity anomaly correction. Geophys J Int 174:1–13
Cheng MK, Tapley BD (2004) Variations in the Earth's oblateness during the past 28 years. J Geophys Res Solid Earth 109(B09402):9. doi:10.1029/2004jb003028
Christensen NI (1982) Seismic velocities. In: Carmichael RS (ed) Handbook of physical properties of rocks, vol 2, pp 1–228. CRC Press, Boca Raton, FL
Chulick GS, Detweiler S, Mooney WD (2013) Seismic structure of the crust and uppermost mantle of South America and surrounding oceanic basins. J S Am Earth Sci 42:260–276
Cordani UG, Sato K (1999) Crustal evolution of the South America Platform based on Nd isotopic systematics of igneous rocks. Episodes 22:167–173
Cooper M (2007) Structural style and hydrocarbon prospectivity in fold and thrust belts: a global review. In: Ries AC, Butler RW, Graham RH (eds) Deformation of the continental crust:

the legacy of Mike Coward (PDF). Special Publications 272. Geological Society, London, pp 447–472. ISBN 978-1-86239-215-1. Retrieved 2 July 2011
Davies GF (1999) Dynamic earth plates. Cambridge University Press, Plumes and Mantle Convection, p 458
De Geer G (1888) Om Skandinaviens nivåförändringar under Quartärperioden. Geol Fören Stockh Förh 10:366–379
Devoti R, Pietrantonio G, Riguzzi F (2014) GNSS networks for geodynamics in Italy. Física de la tierra, no 26, pp 11–24. ISSN 0214-4557
Divins DL (2003) Total sediment thickness of the World's oceans & marginal seas. Boulder, Colo
Dobrin MB (1976) Introduction to geophysical prospecting (3d ed). McGraw-Hill, New York, 630 p
Dorman LM, Lewis TR (1970) Experimental Isostasy: 2. An isostatic model for the U.S.A. derived from gravity and topographic data. J Geophys Res 75, p 3367
Dziewonski AM, Anderson DL (1981) Preliminary reference Earth model. Phys Earth Planet Inter 25:297–356
Ebbing J, England RW, Korja T, Lauritsen T, Olesen O, Stratford W, Weidle C (2012) Structure of the Scandes lithosphere from surface to depth. Tectonophysics 536–537:1–24. 16 Apr 2012
Eken T, Shomali ZH, Roberts R, Hieronymus CF, Bodvarsson R (2008) S and P velocity heterogeneities within the upper mantle below the Baltic Shield. Tectonophysics 462:109–124. doi:10.1016/j.tecto.2008.02.015
Ekman M (1991) Course and origin of the Fennoscandian uplift: the case for two separate mechanisms. Terra Nova 3:408–413
Ekman M, Mäkinen J (1996) Recent postglacial rebound, gravity change and mantle flow in Fennoscandia. Geophys J Int 126(1):229–234. doi:10.1111/j.1365-246X.1996.tb05281.x
Farrell WE (1972) Deformation of the Earth by surface loading. Rev Geophys 10:761–797
Fjeldskaar W, Cathles L (1991) The present rate of uplift of Fennoscandia implies a low-viscosity asthenosphere. Terra Nova 3:393–400
Forte AM, Mitrovica JX (1996) New inferences of mantle viscosity from joint inversion of long-wavelength mantle convection and post-glacial rebound data. Geophys Res Lett 23(10): 1147–1150
Fraeijs de Veubeke BM (1979) A course in elasticity. Springer, New York
Gerhard Franz, Lucassen Friedrich, Kramer Wolfgang, Trumbull Robert B, Romer Rolf L, Wilke Hans-Gerhard, Viramonte José G, Becchio Raúl, Siebel Wolfgang (2006) The Andes, crustal evolution at the Central Andean continental margin: a geochemical record of crustal growth, recycling and destruction. Front Earth Sci 2006:45–64
Gorbatchev R (2004) The Transscandinavian IGNEOUS belt—introduction and background. In: Högdahl K, Andersson U, Eklund O (eds) The Transscandinavian Igneous Belt (TIB) in Sweden: a review of its character and evolution. Geological survey of Finland, Special Paper vol 37. Geological Survey of Finland, pp 9–15
Goiginger H, Rieser D, Mayer-Guerr T, Pail R, Schuh W-D, Jäggi A, Maier A (2011) GOCO consortium, the combined satellite-only global gravity field model GOCO02S. European Geosciences Union General Assembly 2011, Wien, 04.04.2011
Grad M, Tiira T, ESC Working Group (2009) The Moho depth map of the European plate. Geophys J Int 176:279–292. doi:10.1111/j.1365-246X.2008.03919.x
Gradmann S, Ebbing J, Fullea J (2013) Integrated geophysical modelling of a lateral transition zone in the lithospheric mantle under Norway and Sweden. Geophys J Int 194:1358–1373
Goslin J, Beuzart P, Frauchetau J, Le Piehon X (1972) Thickening of the oceanic layer in the Pacific Ocean. Mar Geophys Res 1:418
Haas R, Scherneck H-G, Gueguen E, Nothnagel A, Campbell J (2002) Large-scale strain-rates in Europe derived from observations in the European geodetic VLBI network. European Geosciences Union 2002 (EGU) Stephan Mueller Special Publication Series, vol 2, pp 139–152
Hackney RI, Featherstone WE (2003) Geodetic versus geophysical perspectives of the 'gravity anomaly'. Geophys J Int 154(35–43):2003

Han S-C, Shum CK, Jekeli C, Kuo C-Y, Wilson C, Seo K-W (2005) Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. Geophys J Int 163:18–25. doi:10.1111/j.1365-246X.2005.02756

Hanssen RF (2001) Radar interferometry: data interpretation and error analysis. Kluwer Academic, ISBN 9780792369455.

Haskell NA (1935) The motion of a fluid under a surface load, 1. Physics 6(265–269):1935

Heidbach O, Tingay M, Barth A, Reinecker J, Kurfeß D, Müller B (2008) The world stress map database release 2008. doi:10.1594/GFZ.WSM.Rel2008

Heiskanen WA, Vening Meinesz FA (1958) The Earth and its gravity field. McGraw-Hill Book Company, Inc, New York

Heiskanen WA, Moritz H (1967) Physical geodesy. W.H. Freeman, New York

Henkel H, Eriksson L (1987) Regional aeromagnetic and gravity studies in Scandinavia. Precambrian Res 35:169–180

Hinze WJ, Aiken C, Brozena J, Coakley B, Dater D, Flanagan G, Forsberg R, Hildenbrand T, Keller GR, Kellogg J (2005) New standards for reducing gravity data: The North American gravity database. Geophysics 70(4):J25–J32

Holmes A (1931) Radioactivity and earth movements. Geol Soc Glasgow Trans 18(559–606): 1931

Horemuz M (2010) Deformation measurements and analysis. Lecture notes. Royal Institute of Technology, Stockholm, School of Architecture and Built Environment, Division of Geodesy and Geoinformatics, Stockholm, October 2010

Hotine M (1969) Mathematical geodesy, ESSA monograph 2. Department of Commerce, Washington

Ilk KH (1983) Ein eitrag zur Dynamik ausgedehnter Körper-Gravitationswechselwirkung. Deutsche Geodätische Kommission, Reihe C, Heft Nr. 288, München

Jacob T, Wahr J, Gross R, Swenson S, Geruo A (2012) Estimating geoid height change in North America: Past, present and future. J Geodesy 86:337–358. doi:10.1007/s00190-011-0522-7

Kaban MK, Schwintzer P, Artemieva IM, Mooney WD (2003) Density of the continental roots: compositional and thermal contributions. Earth Planet Sci Lett 209:53–69

Kaban MK, Schwintzer P, Reigber Ch (2004) A new isostatic model of the lithosphere and gravity field. J Geodesy 78:368–385

Kaban MK, Tesauro M, Cloetingh S (2010) An integrated gravity model for Europe's crust and upper mantle. Earth Planet Sci Lett 296(3–4):195–209. ISSN 0012-821X, doi:10.1016/j.epsl.2010.04.041 (1 Aug 2010)

Kaufmann G, Lambeck K (2000) Mantle dynamics, postglacial rebound and the radial viscosity profile. Phys Earth Planet Inter 121:301–324

Kaufmann G, Wu P, Guoying L (2000) Glacial isostatic adjustment in Fennoscandia for a laterally homogeneous earth. Geophys J Int 143:262–273

Kennett BLN, Engdahl ER (1991) Travel times for global earthquake location and phase identification. Geophys J Int 105:429–465

Kierulf HP, Steffen H, Simpson MJR, Lidberg M, Wu P, Wang H (2014) A GPS velocity field for Fennoscandia and a consistent comparison to glacial isostatic adjustment models. J Geophys Res Solid Earth 119(8):6613–6629

King Scott D (1995) Models of mantle viscosity. Sciences, West Lafayette, IN 47907-1397. Mineral Physics and Crystallography A Handbook of Physical Constants. AGU Reference Shelf 2

Klees R, Revtova EA, Gunter BC, Ditmar P, Oudman E, Winsemius HC, Savenije HHG (2008) The design of an optimal filter for monthly GRACE gravity models. Geophys J Int 175: 417–432. doi:10.1111/j.1365-246X.2008.03922.x

Kolstrup ML, Pascal C, Maupin V (2012) What compensates the topography of southern Norway? Insights from thermo-isostatic modelling. J Geodynamics 61:105–119

Korja A (1995). Structure of the Svecofennian crust growth and destruction of the Svecofennian Orogen (Dissertation). Institute of Seismology, University of Helsinki, Report S-31, 36 pp

Korja A, Korja T, Luosto U, Heikkinen P (1993) Seismic and geoelectric evidence for collisional and extensional events in the Fennoscandian shield—implications for precambrian crustal evolution. In: Green AG, Kröner A, Götze J-J, Pavlenkova N (eds) Plate Tectonic signatures in the continental lithosphere. Tectonophysics, vol 219, pp 129–152

Kroll H, Kirfel A, Heinemann R, Barbier B (2012) Volume thermal expansion and related thermophysical parameters in the Mg, Fe olivine solid-solution series. Eur J Miner 24:935–956

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. J Geodesy 81(11):733–749

Kusche J, Schmidt R, Petrovic S, Rietbroek R (2009) Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. J Geodesy. doi:10.1007/s00190-009-0308-3

Lambeck K, Johnston P, Smither C, Nakada M (1996) Glacial rebound of the British Isles. III. Constraints on mantle viscosity. Geophys J Int 125:340–354

Lambeck K, Smither C, Johnston P (1998) Sea-level change, glacial rebound and mantle viscosity for northern Europe. Geophys J Int 134:102–144

Laske G, Masters G, Reif C (2000) A new global crustal model at 2×2 degrees (CRUST2.0). (http://igppweb.ucsd.edu/~gabi/rem.dir/crust/crust2.html)

Laske G, Masters G, Ma Z, Pasyanos M (2013) Update on CRUST1.0—A 1-degree global model of Earth's crust. Geophys Res Abstracts, vol 15, Abstract EGU2013-2658

Lemoine JM, Bruinsma S, Loyer S, Biancale R, Marty J-C, Perosanz F, Balmino G (2007) Temporal gravity field models inferred from GRACE data. Adv Space Res 39(1620–1629):2007

Le Pichon X (1968) Sea-floor spreading and continental drift. J Geophys Res 73:3661–3697

Li X, Götze H-J (2001) Tutorial: ellipsoid, geoid, gravity, geodesy, and geophysics. Geophysics 66:1660–1668 (with the Erratum in 67, 997)

Lidberg M, Johansson JM, Scherneck H-G, Davis JL (2007) An improved and extended GPS derived velocity field for the glacial isostatic adjustment in Fennoscandia. J Geodesy 81(3):213–230. doi:10.1007/s00190-006-0102-4

Lidberg M, Johansson JM, Scherneck H-G, Milne GA (2010) Recent results based on continuous GPS observations of the GIA process in Fennoscandia from BIFROST. J Geodyn 50:8–18. doi:10.1016/j.jog.2009.11.010

Lloyd S, van der Lee S, França GS, Assumpção M, Feng M (2010) Moho map of South America from receiver functions and surface waves. J Geophys Res 115:B11315. doi:10.1029/2009JB006829

Liu HS (1977) Convection pattern and stress system under the African plate. Phys Earth Planet Int 15:60–68

Lowrie W (2007) Fundamentals of geophysics. Cambridge University Press. The Edinburgh Building, Cambridge CB2 8RU, UK. ISBN-13 978-0-511-35447-2

Luosto U (1997) Structure of the Earth's crust in Fennoscandia as revealed from refraction and wide-angle reflection studies. Geophysica 33(1):3–16

Makhloof A (2007) The use of topographic-isostatic mass information in geodetic applications, Doctoral dissertation. Department of Theoretical and Physical Geodesy, Bonn, Germany

Massonnet D, Feigl KL (1998) Radar interferometry and its application to changes in the earth's surface. Rev Geophys 36(4):441–500. doi:10.1029/97RG03139

Mayer-Guerr T, Rieser D, Höck E, Brockmann JM, Schuh, W-D, Krasbutter I, Kusche J, Maier A, Krauss S, Hausleitner W, Baur O, Jäggi A, Meyer U, Prange L, Pail R, Fecher T, Gruber T (2012) The new combined satellite only model GOCO03s. Abstract, GGHS2012, Venice

Martinec Z (1994) The minimum depth of compensation of topographic masses. Geophys J Int 117:545–554

Maupin V, Agostini A, Artemieva I, Balling N, Beekman F, Ebbing J, England RW, Frassett A, Gradmann S, Jacobsen BH, Köhler A, Kvarven T, Medhus AB, Mjelde R, Ritter J, Sokoutis D, Stratford W, Thybo H, Wawerzinek B, Weidle C (2013) The deep structure of the Scandes and its relation to tectonic history and present-day topography. Tectonophysics. doi:10.1016/j.tecto.2013.03.010

McConnell RK (1965) Isostatic adjustment in a layered earth. J Geophys Res 70:5171
McKenzie DP, Parker RL (1967) The North Pacific: an example of tectonics on a sphere. Nature 216:1276–1280
McKenzie D (1978) Some remarks on the development of sedimentary basins. Earth Planet Sci Lett 40:25–32
Medhus AB, Balling N, Jacobsen BH, Weidle C, England RW, Kind R, Thybo H, Voss P (2012) Upper-mantle structure beneath the Southern Scandes Mountains and the Northern Tornquist Zone revealed by P-wave travel time tomography. Geophys J Int 189(3):1315–1334
Meissner R (1973) The Moho as a transition zone. Geophys Surv 1(195–216):1973
Meissner R, Wever Th, Fluh ER (1987) The Moho in Europe—implications for crustal development. Ann Geophys 5B(4):357
McKenzie D, Jackson J, Priestley K (2005) Thermal structure of oceanic and continental lithosphere. Earth Planet Sci Lett 233:337–349
Milne GA, Davis JL, Mitrivica JX, Scherneck H-G, Johansson JM, Vermeer M, Koivula H (2001) Space geodetic constraints on glacial isostatic adjustment in Fennoscandia. Science 291: 2381–2385
Mishra DC, Ravi Kumar M (2012) Long and short wavelengths of Indian Ocean geoid and gravity lows: mid-to-upper mantle sources, rapid drift and seismicity of Kachchh and Shillong plateau, India. J Asian Earth Sci 60:212–224. 22 Oct 2012
Moritz H (1990) The figure of the Earth. Wichmann H, Karlsruhe
Moritz H (2000) Geodetic reference system 1980. J Geod 74(1):128–162. doi:10.1007/S001900050278
Mooney WD, Laske G, Masters TG (1998) CRUST 5.1: a global crustal model at 5 × 5 deg. J Geophys Res 103:727–747
Moore P, Zhang Q, Alothman A (2005) Annual and semiannual variations of the Earth's gravitational field from satellite laser ranging and CHAMP. J Geophys Res 110:2005. doi:10.1029/2004JB003448
Müller RD, Sdrolias M, Gaina C, Roest WR (2008) Age, spreading rates, and spreading asymmetry of the world's ocean crust. Geochem Geophys Geosyst 9:Q04006
Nansen F (1928) The Earth's crust, its surface forms and isostatic adjustment, Norske Videnskaps-Akad. Oslo, I. Mat.-Naturv. Kl., Vol, Avhandl, p 12
Nicolls RJ, Cazenave A (2010) Sea-level rise and its impact on coastal zones. Science 328: 1517–1520
Nieuwland DA (ed) (2003) New insights into structural interpretation and modelling, p 101. Geological Society, London, Special Publications, vol 212, pp 101–116. 0305-8719/03/$15 © The Geological Society of London 2003
Niskanen E (1949) On the elastic resistance of the earth's crust. Ann Acad Sci Fenn, AIII 21:1–23
Noomen R, Springer TA, Ambrosius BAC, Herzberger K, Kuijper DC, Mets G-J, Overgaauw B, Wakker KF (1996) Crustal deformations in the Mediterranean area computed from SLR and GPS observations. J Geodyn 21:73–96
Oldenburg DW (1974) The inversion and interpretation of gravity anomalies. Geophysics 39: 526–536
Ortlieb L, Ruegg JC, Angelier J, Colletta B, Kasser M, Lesage P (1989) Geodetic and tectonic analyses along an active plate boundary: the central Gulf of California. Tectonics 8(3):429–441. doi:10.1029/TC008i003p00429
Pavlis N, Factor K, Holmes SA (2007) Terrain-Related Gravimetric Quantities Computed for the Next EGM. Presented at the 1st international symposium of the international gravity service 2006, August 28–September 1, Istanbul, Turkey
Pavlis N, Holmes SA, Kenyon SC, Factor JK (2008) An Earth gravitational model to degree 2160: EGM08. Presented at the 2008 general assembly of the European Geosciences Union, Vienna, Austria, 13–18 Apr 2008
Pavlis NK, Holmes SA, Kenyon SC, Factor JK (2012) The development and evaluation of the Earth gravitational model 2008 (EGM2008). J Geophys Res 117:B04406

Parsons B, Richter FM (1980) A relation between the driving force and geoid anomaly associated with mid-ocean ridges. Earth Planet Sci Lett 51:445–450. doi:10.1016/0012-821X(80)90223-X
Parker RL (1972) The rapid calculation of potential anomalies. Geophys J R Astr Soc 31:447–455
Peltier WR (1974) The impulse response of a Maxwell Earth. Rev Geophys Space Phys 12: 649–669
Peltier WR, Andrews JT (1976) Glacial isostatic adjustment I. The forward Problem. Geophys J R Astron Soc 46:605–646
Peltier WR (2004) Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G(VM2) model and GRACE. Annu Rev Earth Planet Sci 32:111–149. doi:10.1146/annurev.earth.32.082503.144359
Peltier WR (2007) Mantle dynamics and the D-double prime layer implications of the post-perovskite phase. In: Hirose K, Brodholt J, Lay T, Yuen D (eds) Post-perovskite: the last mantle phase transition, vol 174. AGU Geophysical Monograph, American Geophysics Union, pp 217–227
Petrovskaya MS, Vershkov AN (2006) Non-singular expressions for the gravity gradients in the local north-oriented and orbital reference frames. J Geod 80:117–127
Pick M (1994) The geoid and tectonic forces. In: Vanicek P, Christou N (eds) Geoid and its geophysical interpretations. CRC Press 386 p
Pratt JH (1855) On the attraction of the Himalaya Mountains and of the elevated regions beyond upon the plumb-line in India. Trans Roy Soc (London), ser B, 145
Prutkin I, Saleh M (2009) Gravity and magnetic data inversion for 3D topography of the Moho discontinuity in the northern Red Sea area, Egypt. J Geodyn 47(5):237–245
Reguzzoni M, Sampietro D (2012) Moho estimation using GOCE data: a numerical simulation. In: Kenyon SC, Pacino MC, Marti U (eds) International Association of Geodesy symposia, Geodesy for planet Earth, vol 136, pp 205–214
Reguzzoni M, Sampietro D, Sansò F (2013) Global Moho from the combination of the CRUST2.0 model and GOCE data. Geophys J Int 195(1):222–237. doi:10.1093/gji/ggt247
Richter FJ, Parsons B (1975) On the interaction of two scales of convection in the mantle. J Geophys Res, Dawson, J.B., 1970. The structural setting of African kimber 80(17):2529
Riguzzi F, Crespi M, Devoti R, Doglioni C, Pietrantonio G, Pisani AR (2013) Strain rate relaxation of normal and thrust faults in Italy. Geophys J Int (2013) 195:815–820. doi:10.1093/gji/ggt304
Rummel R, Rapp RH, Sünkel H, Tscherning CC (1988) Comparisons of global topographic isostatic models to the Earth's observed gravity field. Report No. 388, Department of Geodetic Science and Surveying, Ohio State University, Columbus
Runcorn S (1962a) Towards a theory of continental drift. Nature 193:311–314
Runcorn S (1962b) Convection currents in the Earth's mantle. Nature 195:1248–1249
Runcorn S (1967) Flow in the mantle inferred from the low degree harmonics of the geopotential. Geophys J Int 14(1–4):375–384. doi:10.1111/j.1365-246X.1967.tb06253.x
Sampietro D (2009) An inverse gravimetric problem with GOCE data, presented in geodesy for planet earth IAG scientific assembly 2009, August 31–Sptember 4, 2009 Buenos Aires, Argentina
Sampietro D, Reguzzoni M, Braitenberg C (2013) The GOCE estimated Moho beneath the Tibetan Plateau and Himalaya. In: Rizos C, Willis P (eds) International Association of Geodesy Symposia, Earth on the edge: science for a sustainable planet, Proceedings of the IAG General assembly, 28 June–2 July 2011, Melbourne, Australia, 139, Springer, Berlin
Sasgen I, Detlef W, Zdenˇek M, Volker K, Jan H (2005) Geodetic signatures of glacial changes in Antarctica: rates of geoid-height change and radial displacement due to present and past ice-mass variations, GFZ, Scientific Technical Report STR05/01
Sauramo M (1958) Die Geschichte der Ostsee. Ann Acad Sci Fenn, AIII 51:1–522
Seton M, Müller RD, Zahirovic S, Gaina C, Torsvik T, Shephard G, Talsma A, Gurnis M, Turner M, Maus S, Chandler M (2012) Global continental and ocean basin reconstructions since 200 Ma. Earth Sci Rev 113:212–270

Scherneck H-G, Lidberg M, Haas R, Johansson JM, Milne GA (2010) Fennoscandian strain rates from BIFROST GPS: a gravitating, thick-plate approach. J Geodyn 50(1):19–26
Schowalter WR (1978) Mechanics of non-newtonian fluids pergamon. ISBN 0-08-021778-8
Sjöberg LE (1975) On the discrete boundary value problem of physical geodesy with harmonic reduction to an internal sphere. Royal Institute of Technology, Division of Geodesy, Stockholm. (Ph.D. Thesis.)
Sjöberg LE (1983) Land uplift and its implications on the geoid in Fennoscandia. Tectonophysics 97:97–101
Sjöberg LE, Nord T, Fan H (1991) The Fennoscandian geoid bulge and its correlation with land uplift and Moho depth. In: Schultz BE, Anderson A, Froidevaux G, Park M (eds) Gravimetry and space techniques applied to geodynamics and ocean dynamics. AGU geophysical monograph series, vol 82, pp 133–142
Sjöberg LE (1994) The total terrain effect in the modified Stokes' formula. In: Sünkel H, Marson I (eds) Gravity and geoid, IAG symposia, vol 113, Springer, pp 616–623
Sjöberg LE, Fan H, Nord T (1994) Further studies on the Fennoscandian gravity field versus the Moho depth and land uplift. Bull. Geodesique 69:32–42
Sjöberg LE (1997) The total terrain effect in gravimetric geoid determinations. Bolletino di Geodesia e Scienze Affini, LVI, pp 209–222
Sjöberg LE (1998a) The exterior Airy/Heiskanen topographic-isostatic gravity potential anomaly and the effect of analytical continuation in Stokes' formula. J Geodesy 72:654–662
Sjöberg LE (1998b) On the Pratt and Airy models of isostatic geoid undulations. J Geodyn 26 (1):137–147
Sjöberg LE (2000) Topographic effects by the Stokes-Helmert method of geoid and quasigeoid determinations. J. of Geodesy 74(2):255–268
Sjöberg LE (2007) Topographic bias by analytical continuation in physical geodesy. J Geod 81:345–350
Sjöberg LE (2009) Solving Vening Meinesz-Moritz inverse problem in Isostasy. Geophys J Int 179(3):1527–1536. doi:10.1111/j.1365-246X.2009.04397.x
Sjöberg LE, Bagherbandi M (2011) A method of estimating the moho density contrast with a tentative application by EGM08 and CRUST2.0. Acta Geophysica 58:1–24. doi:10.2478/s11600-011-0003-7
Sjöberg LE (2013) On the isostatic gravity anomaly and disturbance and their applications to Vening Meinesz-Moritz inverse problem of isostasy. Geophys J Int 193:1277–1282
Sjöberg LE, Bagherbandi M (2013) A study on the Fennoscandian post-glacial rebound as observed by present-day uplift rates and gravity field model GOCO02S. Acta Geod Geophys 48:317–331. doi:10.1007/s40328-013-0025-5
Sjöberg LE, Bagherbandi M, Tenzer R (2015) On gravity inversion by no-topography and rigorous isostatic gravity anomalies. Pure Appl Geophys 172(10):2669–2680. doi:10.1007/s00024-015-1032-y
Shin YH, Shum CK, Braitenberg C, Lee SM, Xu H, Choi KS, Baek JH, Park JU (2009) Three-dimensional fold structure of the Tibetan Moho from GRACE gravity data. Geophys Res Lett 36(1)
Skinner BJ, Porter SC (1995) The blue planet. An introduction to earth system science, xxiii + 493 + lii pp. Wiley, New York. Price £47.95 (hard covers). ISBN 0 471 54021 8
Spada G, Barletta VR, Klemann V, Riva REM, Martinec Z, Gasperini P, Lund B, Wolf D, Vermeersen LLA, King MA (2011) Benchmark study for glacial isostatic adjustment codes. Geophys J Int 185:106–132
Steffen H, Wu P, Wang H (2010) Determination of the Earth's structure in Fennoscandia from GRACE and implications for the optimal post-processing of GRACE data. Geophys J Int 182:1295–1310
Steffen H, Wu P, Wang H (2012) Optimal locations for absolute gravity measurements and sensitivity of GRACE observations for constraining glacial isostatic adjustment on the northern hemisphere. Geophys J Int 190(1483–1494):2012. doi:10.1111/j.1365-246X.2012.05563.x

Sun W, Sjöberg LE (1999) Gravitational potential changes of a spherically symmetric earth model caused by a surface load. Geophys J Int 137:449–468
Swenson S, Wahr J (2006) Post-processing removal of correlated errors in GRACE data. Geophys Res Lett 33:L08402. doi:10.1029/2005GL025285
Swenson S, Chambers D, Wahr J (2008) Estimating geocentre variations from a combination of GRACE and ocean model output. J Geophys Res Solid Earth 113(B08410):12. doi:10.1029/2007jb005338
Swieczak M, Kozlovskaya E, Majdanski M, Grad M (2009) Interpretation of geoid anomalies in the contact zone between the East European Craton and the Paleozoic platform—II: modeling of density in the lithospheric mantle. Geophys J Int 177:334–346
Sünkel H (1985) An isostatic earth model. In: The department of geodetic science and surveying report no. 367. The Ohio State University, Columbus
Talebian M, Jackson J (2004) A reappraisal of earthquake focal mechanisms and active shortening in the Zagros mountains of Iran (PDF). Geophys J Int (Royal Astronomical Society) 156 (3):506–526. Bibcode:2004GeoJI.156..506T. doi:10.1111/j.1365-246X.2004.02092.x. Retrieved 2 July 2011
Tenzer R, Novák P, Vajda P, Gladkikh V, Hamayun (2012) Spectral harmonic analysis and synthesis of Earth's crust gravity field. Comp Geosci 16(1):193–207
Tenzer R, Hamayun VP (2009) A global correlation of the step-wise consolidated crust-stripped gravity field quantities with the topography, bathymetry, and the CRUST2.0 Moho boundary. Contributions Geophys Geodesy 39(2)133–147
Tenzer R, Bagherbandi M (2012) Reformulation of the Vening–Meinesz Moritz inverse problem of isostasy for isostatic gravity disturbances. Int J Geosci 2012(3):918–929. doi:10.4236/ijg.2012.325094
Tenzer R, Wenjin C, Tsoulis D, Bagherbandi M, Sjöberg LE, Novák P (2014) Analysis of the refined CRUST1.0 crustal model and its gravity field. Surveys in Geophysics 36:139–165. doi:10.1007/s10712-12-014-9299-6
Tesauro M, Kaban MK, Cloetingh S (2008) EuCRUST-07: a new reference model for the European crust. Geoph Res Let 35. doi:10.1029/2007GL032244
Turcotte D, Schubert G (2002) Geodynamics, 2nd ed., Cambridge University Press
Vajda P, Vaníček P, Novak P, Tenzer R, Ellmann A (2007) Secondary indirect effects in gravity anomaly data inversion or interpretation. J Geophys Res 112(B06411)
Vening Meinesz FA (1931) Une nouvelle methode pour la reduction isostatique regionale del'intensite de la pesanteur. Bull Geod 29:33–51
Walcott RI (1980) Rheological models and observational data of glacio-isostatic rebound. In: Mörner N-A (ed) Earth rheology, isostasy and eustasy. Wiley, Chichester
Wallace K, Yin G, Bilham R (2004) Inescapable slow slip on the Altyn Tagh fault. Geophys Res Lett 31(9):1–4. Bibcode: 2004GeoRL. 3109613W. doi:10.1029/2004GL019724
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. J Geophys Res 103, 30205–30229
Watts AB (2001) Isostasy and flexure of the lithosphere. Cambridge University Press, Cambridge
Wegener A (1912) Die Entstehung der Kontinente. Sid. 185–195(253–256):305–309
Wegener A (1924) The origin of continents and oceans. Methuen and Co., Ltd, p 1924
Wild F, Heck B (2004a) A comparison of different isostatic models applied to satellite gravity gradiometry, Gravity, Geoid and Space Missions GGSM 2004 IAG International Symposium Porto, Portugal August 30–September 3
Wild F, Heck B (2004b) Effects of topographic and isostatic masses in satellite gravity gradiometry. In: Proceedings of the second international GOCE user workshop GOCE. The Geoid and Oceanography, ESA-ESRIN, Frascati/Italy, 8–10 Mar 2004 (ESA SP–569, June 2004), CDROM
Wouters B, Schrama EJO (2007) Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering of spherical harmonics. Geophys Res Lett 34:L23711. doi:10.1029/2007GL032098

Wu P, Steffen H, Wang HS (2010) Optimal locations for GPS measurements in North America and northern Europe for constraining Glacial Isostatic Adjustment. Geophys J Int 181(653–664):2010. doi:10.1111/j.1365-246X.2010.04545.x

Zhang P-Z, Molnar P, Xu X (2007) Late quaternary and present-day rates of slip along the Altyn Tagh fault, northern margin of the Tibetan Plateau. Tectonics 27 (TC5010): 1–24. Bibcode:2007Tecto..26.5010Z. doi:10.1029/2006TC002014

Zhao S, Lambeck K, Lidberg M (2012) Lithosphere thickness and mantle viscosity inverted from GPS-derived deformation rates in Fennoscandia. Geophys J Int 190:278–292

Zhong S, McNamara A, Tan E, Moresi L, Gurnis M (2008) A benchmark study on mantle convection in a 3-D spherical shell using CitcomS. Geochem Geophys Geosyst 9(10). doi:10.1029/2008GC002048. http://maps.unomaha.edu/Maher/plate/week13/mechanisms.html

Chapter 9
Concluding Remarks and Outlook

Extensive knowledge of the Earth's gravity field is most versatile in geomatic/technical applications in society, as well as in geosciences. The increasing use of geospatial data calls for precise reference systems, as acknowledged by a United Nations' resolution in 2015 (see Sect. 1.5). In this context, a precise geoid model is needed as the vertical zero level in a 3-D coordinate system, and this issue is particularly demanding in applications of precise satellite positioning, such as GNSS.

The last 30–40 years have been a very intensive period of development in physical geodesy and gravimetric geophysics. This rapid development was partly fuelled by the need for accurate geoid models in the new technology of GNSS surveying, but the progress was also encouraged by scientific needs in geophysics (e.g. oceanography) and climate change. These needs have now (in 2016) seen three successful dedicated satellite-gravity missions being realized in the first decade of the new millennium, and the outcome of these efforts, e.g. for geoscience studies of planet Earth and climate change, have been far beyond expectations. We have also seen a fast development of EGMs, suitable for global and regional gravity-field and geoid modelling. Since the 1990s, the 1-cm geoid has been a primary goal, which has now been met in a few countries with smooth topography and/or high-quality gravity networks. In the future, even higher-quality geoid models are likely to be requested, and their realizations will put further demands on both theory and data. The LSMSA technique for geoid determination may possibly be improved by adding a Meissl type of modification in the least-squares modification of Stokes' formula, with extended integration of cap size. Gravimetric data will improve both in quality and coverage at all wavelengths. For covering large areas with rather detailed data, airborne gravimetry has become an important tool, and for determining the global long-wavelength gravity field a GRACE Follow-On mission is planned, using again satellite-to-satellite tracking for inter-satellite range and range-rate determination by microwave interferometry, with at least the same accuracy as in the first mission, but also laser interferometry will be tested for

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4_9

higher accuracy. In due time, one can also foresee a follow-up of the successful GOCE satellite gradiometry mission.

In the future, vertical control points, established, e.g. by steel bench marks in hard rock, may possibly be abandoned in favour of vertical control systems based on high-quality (digital) geoid models, suitable for height determination by GNSS-levelling. Today, most European and former Soviet countries define their vertical systems in terms of normal height and quasigeoid, while most other countries use the orthometric height and the geoid in their height systems. [Some comparisons are given in Vanicek et al. (2012) and in Sjöberg (2013)]. The primary advantage of the quasigeoid-based system is that the partly unknown mass distribution of the topography does not affect the computed (quasigeoid and normal) heights and their accuracies. On the other hand, the geoid is a natural zero level as it approximates mean sea level (and its prolongation through the continents), while the quasigeoid has no such natural interpretation. In addition, the geoid, in contrast to the quasigeoid, is an equipotential surface in the Earth's gravity field. Therefore, the geoid, but not the quasigeoid (which is a projection of the telluroid that closely follows the mostly irregular topographic surface), can be successfully digitized and interpolated to very high resolution, and this fact could possibly be a practical problem in future quasigeoid-based systems.

While geoid computation is successfully achieved by combining the long- to medium-wavelength data of an EGM with a Stokes' type integral, which provides the detailed information, an EGM alone limited to, say, 180° is sufficient to cover all significant wavelengths in some geophysical applications, such as gravity inversion for Moho depth and (possibly) density contrast, as well as for tectonic stress in the mantle. For such studies, an improved low- to medium-wavelength EGM will be beneficial.

Among the remaining problems to be solved is the bias in gravity field data over the oceans caused by the Sea Surface Topography in the process of inverting satellite-altimetry data to gravity data. This issue is important, as the detailed gravity field on the ocean mainly relies on this technique, and there is by now no successful independent way to remove the bias, of the order of ± 1–2 m, and a considerable part of it therefore remains in the gravity.

Today, the GRACE mission is, and its successor(s) will be, the most essential techniques for determining the global to regional temporal changes of gravity and geoid with several important practical and scientific applications for science and society. Examples of such applications are monitoring of geoid and sea level changes and land uplift rates, as well as mass changes of glaciers, ice sheets and hydrography. A notable application in oceanography is to use the estimated temporal change of slope of the geoid, which, combined with the rate of change of sea level slope, provides the change of steric ocean current, closely related with climate change.

Today the common way to estimate the quality of a gravimetric geoid determination is to compare it with a GNSS-levelling based geometric geoid model. As this comparison includes independent systems for gravity, GNSS and levelling, much of the systematic errors of each system are lumped together and determined in

a least squares fitting procedure and removed, but there is no reliable way to distribute the estimated bias and tilts to gravimetric and GNSS-levelling geoid results. Interestingly, the emerging research towards a relativistic technique by atomic clocks will offer another independent means for height and geoid determinations and control, e.g. in a global network of control points as complementary to satellite gravimetry, as well as to the global geometric reference of the ITRF. In the future, chronometric levelling may, for the first time, even become a practical tool for direct measuring of geopotential differences for various applications in geodesy and geophysics (see, e.g. Kopeikin et al. 2015; Flury 2016).

References

Flury J (2016) Relativistic geodesy. J Phys Conf Ser 723:012051

Kopeikin SM, Mazurova EM, Karpik AP (2015) Towards an exact relativistic theory of Earth's geoid undulation. Phys Lett A 379:1555–1562

Sjöberg LE (2013) The geoid or quasigeoid - which surface should be preferred for a national height system? J Geod Sci 3:103–109

Vanicek P, Kingdon R, Santos M (2012) Geoid versus quasigeoid : a case of physics versus geometry. Contr Geophys Geod 42(1):101–117

Index

L.E. Sjöberg and M. Bagherbandi, *Gravity Inversion and Integration*,
DOI 10.1007/978-3-319-50298-4