Institut für Geodäsie und Geoinformation der Universität Bonn

---

Gravity Field Refinement by Radial Basis Functions from In-situ Satellite Data

Inaugural–Dissertation zur
Erlangung des akademischen Grades
Doktor–Ingenieur (Dr.–Ing.)
der Hohen Landwirtschaftlichen Fakultät
der Rheinischen Friedrich–Wilhelms–Universität
zu Bonn

vorgelegt am 30. Januar 2008 von

Dipl.–Ing. Annette Eicker
aus Haan

D 98

Referent: Prof. Dr.–Ing. K.–H. Ilk

Korreferenten: Prof. Dr. rer. nat. H.–P. Helfrich
Prof. Dr.–Ing. R. Rummel

Tag der mündlichen Prüfung: 28. März 2008

Publikation: Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert. (Erscheinungsjahr 2008)

Gravity Field Refinement by Radial Basis Functions from In-situ Satellite Data

Summary

In this thesis, an integrated approach is developed for the regional refinement of global gravity field solutions. The analysis concepts are tailored to the in-situ type character of the observations provided by the new satellite missions CHAMP, GRACE, and GOCE. They are able to evaluate data derived from short arcs of the satellite's orbit and, therefore, offer the opportunity to use regional satellite data for the calculation of regional gravity field solutions. The regional character of the approach will be realized at various stages of the analysis procedure. The first step is the design of specifically tailored space localizing basis functions. In order to adapt the basis functions to the signal content to be expected in the gravity field solution, they will be derived from the covariance function of the gravitational potential. To use the basis functions in gravity field modeling, they have to be located at the nodal points of a spherical grid; therefore investigations will be performed regarding a suitable choice of such a nodal point distribution. Another important aspect in the regional gravity field analysis approach is the downward continuation process. In this context, a regionally adapted regularization will be introduced which assigns different regularization matrices to geographical areas with varying signal content. Regularization parameters individually determined for each region take into account the varying frequency behavior, allowing to extract additional information out of a given data set. To conclude the analysis chain, an approach will be described that combines regional solutions with global coverage to obtain a global solution and to derive the corresponding spherical harmonic coefficients by means of the Gauss-Legendre quadrature method. The capability of the method will be demonstrated by its successful application to real data provided by CHAMP and GRACE and to a simulation scenario based on a combination of GRACE and GOCE observations.

Verfeinerungen des Gravitationsfeldes mit radialen Basisfunktionen aus in-situ Satellitendaten

Zusammenfassung:

In der vorliegenden Arbeit wird ein ganzheitliches Konzept für die regionale Verfeinerung globaler Gravitationsfeldmodelle entwickelt. Die dazu verwendeten Analyseverfahren sind dem in-situ Charakter der Beobachtungen der neuen Satellitenmissionen CHAMP, GRACE und GOCE angepasst. Sie beruhen auf kurzen Bahnbögen und ermöglichen somit die Berechnung regionaler Gravitationsfeldmodelle aus regional begrenzten Satellitendaten. Der regionale Charakter des Ansatzes wird dabei auf verschiedenen Ebenen des Analyseprozesses realisiert. Der erste Schritt ist die Entwicklung angepasster ortslokalisierender Basisfunktionen. Diese sollen das Frequenzverhalten des zu bestimmenden Gravitationsfeldes widerspiegeln; sie werden daher aus der Kovarianzfunktion des Gravitationspotentials abgeleitet. Um die Basisfunktionen für die Schwerefeldmodellierung zu verwenden, müssen sie an den Knotenpunkten eines sphärischen Gitters angeordnet werden. Daher werden Untersuchungen durchgeführt, welche Punktverteilung für diese Aufgabe besonders geeignet ist. Einen wichtigen Aspekt bei der regionalen Gravitationsfeldanalyse stellt der Fortsetzungsprozess nach unten dar. In diesem Zusammenhang wird ein regional angepasstes Regularisierungsverfahren entwickelt, das verschiedene Regularisierungsmatrizen für regionale Gebiete mit unterschiedlichem Schwerefeldsignal ermöglicht. Individuell angepasste Regularisierungsparameter berücksichtigen den variierenden Signalinhalt, wodurch erreicht wird, dass zusätzliche Informationen aus einem gegebenen Datensatz extrahiert werden können. Schließlich wird ein Ansatz vorgestellt, der regionale Lösungen mit globaler Überdeckung zu einer globalen Lösung zusammenfügt und die zugehörigen sphärischen harmonischen Koeffizienten mit Hilfe der Gauss-Legendre-Quadratur berechnet. Die Leistungsfähigkeit des beschriebenen Ansatzes wird durch eine erfolgreiche Anwendung auf die Echtdatenanalyse aus Daten der Satellitenmissionen CHAMP und GRACE und auf ein Simulationsszenario aus einer Kombination simulierter GRACE– und GOCE–Beobachtungen verdeutlicht.

Contents

1. Introduction

Information about the Earth's gravitational field provides valuable insight into the Earth as a complex dynamic system. Its exact knowledge offers a fundamental contribution to geodetic as well as to a wide variety of geophysical applications. Concerning the geodetic aspects, one important example is the use of the geoid (i.e. an equipotential surface of the Earth's gravity field) as a reference surface for the definition of height systems.To know the location of the geoid exactly is necessary for the derivation of a unified global height datum. As the geoid constitutes the hypothetical ocean surface at rest, it also plays a decisive role in the determination of ocean dynamics. The sea surface can be measured with high accuracy by altimeter satellites, but only by reference to the geoid the ocean circulation can actually be derived from these measurements. Further examples of the impact of the gravity field determination on different geoscientific disciplines can be identified in form of beneficial contributions, e.g. in the areas of hydrology, glaciology or solid Earth's dynamics. Especially the time variable part of the gravity field allows important insights into mass transport and mass variation phenomena. An overview of these processes is given by Ilk et al. (2005b).

The determination of the gravity field of the Earth from satellite data has been performed during the last four decades, but with the launch of the new dedicated gravity field missions a new epoch has started. CHAMP was the first of the new gravity field missions, in orbit since 2000, and has since then provided a new level of gravity field accuracy and valuable experiences being essential for the subsequent missions. GRACE, launched in the year 2002, is the second mission. Besides the determination of the static part of the gravity field in the low and medium wavelengths, GRACE is especially dedicated to the detection of time variable gravity field features. As an example, the temporal and spatial variations of the continental water storage have, for the first time, become observable on a global scale by the evaluation of GRACE observations. The third mission will be GOCE with an anticipated launch date in the early summer of 2008. Its goal is the recovery of the high resolution static part of the gravity field with unprecedented accuracy.

The innovative concept of the new gravity field missions is the in-situ type character of the observations. Each of the satellites carries a GPS receiver, and therefore a continuous observation of the respective satellite's orbit is enabled. Furthermore, the additional observations (precise intersatellite ranges and range-rates in case of GRACE and gradiometer measurements in case of GOCE) allow the gravity field to be exploited more or less directly, and therefore emphasize the in-situ character. In contrast to earlier efforts in gravity field determination from satellite data, the new observation types do not require the analysis of satellite orbits over a long period of time, but are able to cope with short arcs of the satellite's orbit. This is an important aspect that has enabled the determination of regional gravity field solutions from satellite data observed over the particular geographical region.

Despite the outstanding results already obtained from the satellite missions CHAMP and GRACE, it is expected that even more information can be extracted out of the given data. As one opportunity in this context a regional modeling can be identified. The conventionally applied modeling in terms of a spherical harmonic expansion offers an insufficient possibility of representing regional gravity field features. The gravity field in different geographical areas exhibits a strongly varying signal content, especially in the higher frequency part of the gravity field spectrum. These regional characteristics cannot properly be taken into account by basis functions with global support as provided by spherical harmonics. The degree of instability of the gravity field recovery process, which increases with higher frequencies in the gravity field signal, varies with the specific gravity field characteristics in different regions. A global regularization applied to account for the ill-posedness of the downward continuation process causes a global filtering of the observations leading to a mean damping of the global gravity field features. Therefore, it is expected that a regionally tailored regularization process would allow more information to be extracted out of the given signal than would be the case in a global analysis.

It is the goal of this thesis to develop an integrated recovery approach for the calculation of regional gravity field solutions. The regional gravity fields are to be calculated as refinements to global models, accounting for the information that is present in the satellite data in addition to the global model. First of all, the integrated approach includes the design of space localizing basis functions that are adequately tailored to this task and their proper distribution on a reference sphere. But not only the gravity field parameterization is to reflect the regional characteristics, but the whole analysis procedure should be regionally adapted. This implies a functional model based on short arcs of the satellite's orbit on the one hand and a regionally adapted regularization process on the other hand. If desired, regional solutions with global coverage are to be merged to a global solution (parameterized by a spherical harmonic expansion) in a direct stable computation step without losing the details of a regional zoom-in. The described considerations are first investigated theoretically and are finally applied to simulated and real satellite data.

This leads to the following organization of the thesis: In Chapter 2, an overview is given of the regional gravity field determination from the new satellite missions. This includes an introduction of the missions CHAMP, GRACE, and GOCE and the characterization of their respective measurement principles. Furthermore, it is specified how the new observation techniques have led to the development of new, tailored analysis strategies and how these strategies have enabled the recovery of regional gravity field solutions. Different approaches to such a tailored processing of the new satellite data are summarized. Subsequently, a detailed motivation for a regional approach is given, followed by an overview of existing regional modeling techniques.

In Chapter 3, the different tools are introduced which are required to perform a regional analysis. Possibly the most important aspect in this context is the specification of the space localizing basis functions to model the regional solution. In this thesis, basis functions are designed that are derived from the degree variances of the gravitational potential. This implies that the basis functions reflect the frequency spectrum of the gravity field to be modeled. Smoothness assumptions based on the design of the basis functions are investigated within the framework of reproducing kernel Hilbert spaces. As a regional modeling does not only consist of the shape of the basis functions, but also of their location on a reference surface, different spherical grids are specified, and it is investigated which of them is particularly adequate to serve as nodal point pattern for the basis functions.

The definition of the functional models to be used in the analysis of the satellite data is presented in Chapter 4. In this context, different gravity field functionals are expressed in terms of space localizing basis functions, resulting in the specification of the observation equations to be used in the recovery process.

Chapter 5 deals with the solution of the observation equations that are set up in Chapter 4. In this context, particular focus is put on the ill-posedness of the downward continuation process. The possibility of regionally adapting the regularization procedure is pointed out as one of the key benefits of the regional analysis approach.

Chapter 6 describes the opportunity to merge regional solutions with global coverage to obtain a global solution. If required, this global solution can be transformed into a representation by spherical harmonics by means of quadrature methods.

The numerical results derived from the application of the described analysis approaches are presented in Chapter 7. This includes a description of the programming system GROOPS, which was developed in our group in Bonn for gravity field calculations and extended for application in regional analysis. Subsequently, a simulation study concerning different specifications of the space localizing basis functions is given. Finally, the real data analysis of CHAMP and GRACE data and a simulation scenario of a combined GRACE/GOCE analysis is presented.

Chapter 8 provides a summary of the applied techniques and obtained results and gives an outlook on possible further research.

2. Regional Gravity Field Recovery Using New Satellite Missions

The innovative observation principles of the satellite missions CHAMP, GRACE, and (in the near future) GOCE have improved our knowledge of the Earth's gravity field by several orders of magnitude during the last couple of years. To demonstrate the potential of these observation techniques, in the first part of this chapter, these dedicated gravity satellite missions will be introduced. Special emphasis is thereby put on the individual measurement principles featured by the different missions to point out their innovative character. In the second section, it is described in which way the new observation techniques have led to the development of new analysis strategies. They partly differ from the approaches established before the availability of the new missions. The main focus is on the fact that these new analysis approaches have provided the opportunity to use short arcs of the satellite's orbit, tailored especially to the recovery of regional gravity field solutions. The benefit of such regional solutions and their advantages in comparison with the calculation of global solutions is then pointed out in the following section. At the same time, this provides the motivation for the regional investigations that will be described throughout the course of this thesis. It is followed by an overview of regional approaches (using various types of modeling techniques) that have been applied in the analysis of gravitational fields from satellite data so far.

2.1 Overview of the Satellite Missions CHAMP, GRACE, and GOCE

2.1.1 CHAMP

The CHAMP (**CHA**llenging **M**inisatellite **P**ayload) mission was proposed by the GeoForschungsZentrum Potsdam in 1994 in cooperation with the German Aerospace Center (Deutsches Zentrum fuer Luft- und Raumfahrt, DLR) and a consortium of industrial companies from the former East and West Germany (Reigber et al. 1999). The development and manufacturing phase began in January 1997 resulting in the launch of the CHAMP mission on July 15, 2000, from Plesetsk, Russia. It was launched into a near polar and near circular orbit (inclination of $\approx 87°$, eccentricity of ≈ 0.0004) with an original altitude of 454 km. After seven years in orbit, it has descended to an orbit height of approximately 340 km. The given altitude constitutes a compromise between the research interests of the CHAMP mission comprising gravity, magnetism as well as atmosphere and ionosphere. The latter two would have preferred an even higher orbit, and a rather lower orbit would have been more suitable for gravity field research. The satellite moves with a speed of about 28,000 km/h and has a mean period of 1.5 h resulting in approximately 15 revolutions per day.

The satellite has a total mass of 522 kg and features a height of 75 cm and an overall length of 8.333 m including a boom of 4.044 m length. The boom is tailored to the requirements of the observation of the magnetic field. In order to avoid disturbances caused by a magnetic stray field, the magnetometer is located at the end of the boom and, in this way, is kept at a sufficient distance from the satellite.

Concerning the task of gravity field recovery, the satellite itself can be regarded as a sensor. It carries an on-board GPS antenna to enable the determination of the precise orbit by the Global Positioning System. As the satellite's orbit is influenced by gravity field disturbances, the analysis of the orbit data can provide information about the structure of the gravity field. Thus the satellite positions represent the primary observable for the task of gravity field determination. This primary measurement principle is known as satellite-to-satellite tracking in the high-low mode (hl-SST), as the orbit of the low flying CHAMP satellite is determined by the higher-flying GPS satellites. Additionally, CHAMP is equipped with an on-board accelerometer to account for non-gravitational forces acting on the satellite such as atmospheric drag, solar

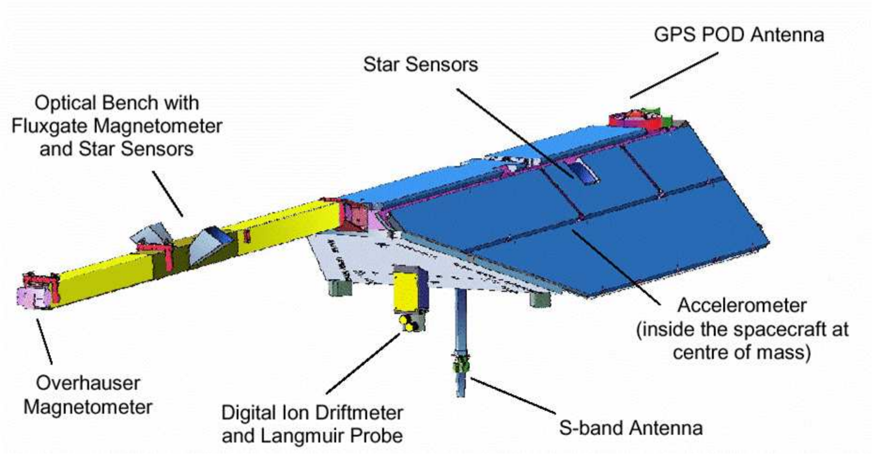


Figure 2.1: Schematic illustration of the CHAMP satellite mission and its key instruments (source: GFZ Potsdam)

radiation, and Earth albedo which influence the orbit as well. Star cameras provide high precision attitude information. Fig. 2.1 shows the most important instruments aboard the satellite in a schematic diagram.

2.1.2 GRACE

GRACE (**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment) is a joint project of the National Aeronautics and Space Administration (NASA) and the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt, DLR). It was proposed in cooperation with the University of Texas in Austin, the Center for Space Research (UTCSR), the GeoForschungsZentrum Potsdam (GFZ), the Jet Propulsion Laboratories (JPL), the Space Systems/Loral (SSL), German Aerospace Center, and the Astrium GmbH in the year 1996. In 1997 GRACE was selected as the second mission in NASA's 'Earth System Science Pathfinder Project' (ESSP). Five years later, on March 17, 2002, the launch took place at Plesetsk, Russia, on a Rockot launch vehicle. A description of the GRACE mission can be found in Tapley et al. (2004). GRACE is a twin satellite mission consisting of two identical satellites following each other in the same orbit separated by a distance of about 220 km. The mission duration was designed to be about five years; at present the lifetime is predicted to be twice as long. Both satellites are equipped with an intersatellite ranging system that establishes the connection by a microwave link enabling the measurement of relative motion (range, range-rate, range-acceleration) between the two satellites with high accuracy. This K-band ranging system is the key instrument of GRACE and is capable of measuring the dual one-way range between both satellites with a precision of about 1 μm. This kind of intersatellite gravity measurement principle is known as low-low satellite-to-satellite tracking (ll-SST). The relative motion is a measure for the inhomogeneities of the gravity field of the Earth, as will be discussed later.

In addition, each satellite carries a GPS receiver to measure its position, thus enabling observations of the type high-low satellite-to-satellite tracking (hl-SST) as well. The satellite configuration displaying the combination of the two satellite-to-satellite tracking measurement principles is illustrated in Fig. 2.2. In addition to this, the onboard accelerometer accounts for non-gravitational forces such as atmospheric drag, solar radiation, and Earth albedo which act on the surface of the satellite and disturb the satellite's orbit.

The altitudes of the two satellites were close to 500 km at the beginning of the mission, but will decline to about 300 km and even lower towards the end of the mission. Therefore, the GRACE twin satellites can be classified as low Earth orbiters (LEOs). The orbit was chosen to be almost circular (with an eccentricity of $e < 0.0005$) and to have an inclination of $89,5°$ to guarantee a homogeneously distributed data coverage.

Besides the determination of a static gravity field solution, the time variable gravity field is a main objective of the GRACE mission.

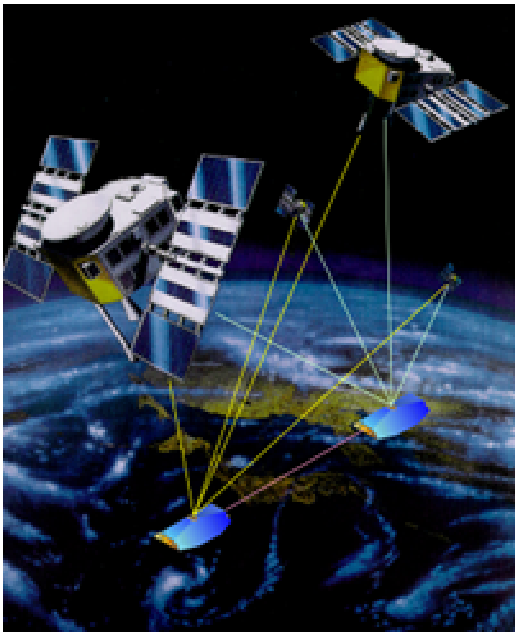


Figure 2.2: Satellite configuration and measurement scenario of the GRACE mission

2.1.3 GOCE

GOCE (**G**ravity Field and Steady-State **O**cean **C**irculation **E**xplorer) is dedicated to measure the Earth's static gravity field and to model the geoid with extremely high accuracy and spatial resolution. It is the first Earth Explorer Core mission to be developed as part of ESA's Living Planet Programme (ESA 1999), and is scheduled for launch in May 2008. A detailed description of the mission can be found, e.g., in DRINKWATER et al. (2003). GOCE is designed to determine the static gravity field of the Earth with unprecedented accuracy and spatial resolution. More specifically, the mission objectives are the determination of the gravity field with a spatial resolution of about 70 km half wavelength and an accuracy of 1 mGal for gravity anomalies and 1-2 cm for the geoid itself. This is possible only because of the extremely low altitude of GOCE of approximately 250 km.

GOCE realizes the concept of satellite gravity gradiometry, i.e. the measurement of differences in gravitational acceleration acting on test masses inside the orbiting satellite, as illustrated in Fig. 2.3. The gradiometer consists of three pairs of orthogonal three-axis accelerometers; each pair of accelerometers is 50 cm apart, with the three axes oriented along-track, across-track, and directed approximately to the Earth's center (as completion of a right-handed triade), respectively. The accelerations acting on each proof mass are determined by measuring the forces necessary to keep the proof mass in the center of its cage. By this constellation, the gradiometer measures continuously all nine components of the gravity gradient tensor, in the practical calculations with different accuracies. Thus the observations provided by GOCE are the second derivatives of the gravitational potential. The gradiometer measurements are supplemented by high-low SST measurements provided by the onboard GPS receiver and by star-tracker information. Ion-thrusters compensate for surface forces acting on the satellite and, in this way, ensure a pure free-fall motion to be permanently maintained.

The satellite will feature an extremely low orbit; its altitude constitutes a compromise between gravity attenuation on the one hand and the influence of the surface forces on the other hand. The inclination will be 96.5°, which results in a lack of observations around the poles, a consequence that is referred to as the polar gap problem. The orbit is designed in a sun synchronous way in order to enable a constant energy supply by the sun panels and to minimize deformation of the satellite caused by temperature fluctuations. The mission profile was designed for energy supply reasons, too. In the original concept, the mission duration was divided into two six-months' measurement phases, in which the time of eclipse is minimal ($<$ 10 min.

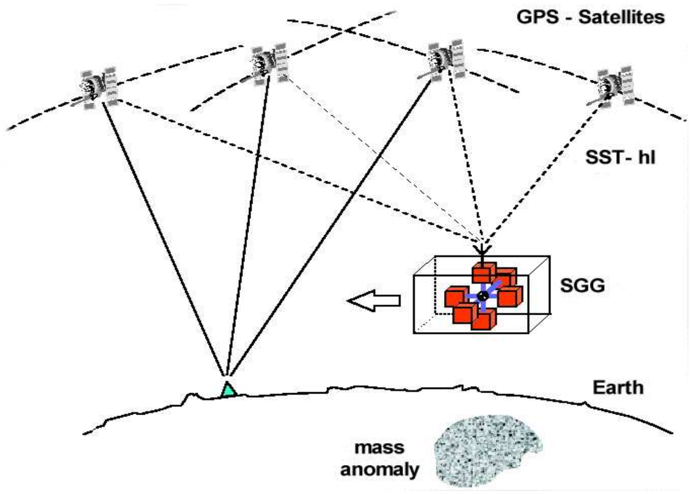


Figure 2.3: GOCE gradiometer measurement principle (source: GOCE Projektbüro Germany)

per orbit), separated by a hibernation phase of about five months during the long-eclipse season. Together with a three months' commission and calibration phase at the beginning of the mission, this resulted in an envisaged overall mission duration of 20 months. Because of the delay of the start of mission, its profile will have to be optimized according to the launch date.

2.2 Regional Gravity Field Recovery

2.2.1 New Analysis Techniques for the New Satellite Missions

Gravity field recovery from satellite data has been performed for the last 40 years. The classical approach of satellite geodesy has thereby been based on the concept of deriving the gravity field parameters from the analysis of accumulated orbit perturbations of artificial satellites with varying orbit altitudes and orbit inclinations. Because of the global characteristics of this technique, the gravitational field has been modeled by a spherical harmonic expansion. Due to the lack of on-board positioning devices such as a GPS receiver in former times, the satellite's positions had to be determined by discrete measurements from the Earth's surface, e.g. by means of precise laser ranging observations. This approach is known as differential orbit improvement technique. Starting from initial values for the satellite's orbit, the partial derivatives of the observations with respect to the unknown parameters are determined by integration of the variational equations. Naturally, when dealing with topocentric observations from only a limited number of observation stations, the coverage with observations is comparably poor. In order to achieve a sufficient redundancy, the use of rather long arcs over days or even weeks was mandatory. In addition to this, the need for long arcs was essential in order to cover the characteristic periodic and secular disturbances caused by the small corrections to the approximated force function parameters. Furthermore, the former gravity field missions were less sensitive to the gravity field signal due to rather high orbits of more than 800 km and featured only limited global coverage due to the given inclinations of the satellites' orbits.

The situation has changed completely with the launch of the present satellite missions CHAMP, GRACE, and (in the near future) GOCE. The innovative character of all three missions is the nearly continuous, precise observation of the respective satellite orbits. In case of GRACE additional, extremely precise range and range-rate K-band measurements are available, and the gradiometer on board the GOCE satellite provides continuous measurements of the gravity gradient. These measurement principles allow the gravity field to be detected more or less directly. Thus, instead of the analysis of accumulated orbit perturbations of artificial satellites caused by the inhomogeneous structure of the gravity field, the new observation techniques can be regarded as having more of an in-situ measurement character. This can easily be understood in the case of satellite gradiometry, as the gradiometer measures directly in-situ the inner geometry of the gravity field, as has been pointed out, e.g., by Rummel (1986). Concerning satellite-to-satellite tracking in the low-low mode, the K-band measurement axis can be regarded as a one-axis gradiometer with a point mass distance of about 200 km. The K-band observations determine the distance between the two satellites at a specific point of time and location; for this reason, they can be understood as in-situ measurements as well. The satellite-to-satellite tracking in the high-low mode with its densely sampled observations can, from its intrinsic characteristics, also be interpreted as being more of an in-situ observation type compared to the classical observation techniques.

These new in-situ type observation principles can be taken into account for the development of new analysis strategies. A variety of different approaches has been proposed and applied to the different types of observations. Ilk et al. (2005a) divide the in-situ measurement principles into three analysis levels. The first among them is based directly on the observed, precisely determined kinematic positions by relating them to the specific force function via an integral equation of Fredholm type. This kind of equation has been used by Mayer-Gürr et al. (2005) for the determination of the CHAMP gravity models ITG-Champ01e, 01k and 01s, and has been adapted to the analysis of GRACE data by Mayer-Gürr (2006). As this analysis concept is also applied to the regional analysis of CHAMP and GRACE data as described in this thesis, it will be introduced in more detail in Sections 4.3.1 and 4.3.2. The second analysis level requires the numerical differentiation of the time series of the positions on the one hand and the integration of the force function on the other hand. This possibility allows to exploit the balance equations of classical theoretical mechanics. The technique has been applied by using the energy balance principle in form of the so-called Jacobi integral; for example Gerlach et al. (2003) used this approach to derive the CHAMP gravity field TUM-1S. Other examples of the application of the energy balance approach can be found, e.g., in Visser et al. (2003), Howe et al. (2003), Kusche and van Loon (2004), and Földvary et al. (2005). Löcher and Ilk (2005) and Löcher and Ilk (2006) investigated additional energy and motion integrals as to their applicability to gravity field recovery and gravity field validation. Among them, there are integrals based on the three scalar components of Newton's equation of motion and integrals based on the linear and angular momentum. According to their investigations, the alternative integrals of motion provide partly better mathematical characteristics than the original Jacobi integral. The third analysis level applies to Newton's equation of motion directly; it thus requires a twofold differentiation of the time series of satellite positions. In an approach introduced by Austen and Reubelt (2000), the satellite positions are approximated by an interpolation polynomial, and by means of a twofold numerical differentiation of the polynomial the local acceleration vector is estimated from the relative GPS position time series. More details concerning the application of this technique to the analysis of CHAMP data can be found in Reubelt et al. (2003) and Reubelt et al. (2006). A similar technique using weighted averages of three successive satellite positions has been applied by Ditmar et al. (2006) for the determination of the gravity model DEOS_CHAMP-01C.

Those analysis concepts tailored to the processing of in-situ observations have in common that they can cope with short arcs of the satellite's orbit. Splitting the orbit into shorter pieces provides the advantage that the accumulated effects of (residual) disturbing forces are kept as low as possible. Furthermore, discontinuities and gaps in the observation series can easily be dealt with by starting a new arc after each data gap. In addition, the use of short arcs provides the opportunity to calculate the gravity field in a specific region from the satellite observations originating from arcs covering the respective region. This is an inevitable requirement for the calculation of regional gravity field solutions. Thus the new observation and analysis procedures open the opportunity of regional gravity field analysis as a method of refining global solutions.

Nevertheless, it has to be mentioned that the classical analysis techniques based on the integration of the variational equations are still being applied very successfully to the new satellite missions as well. An example is the so-called dynamic approach (as described in Schmidt 2007) which has been applied to determine the EIGEN gravity field series (see, for example, Förste et al. 2005). Further examples are the calculation of, e.g., the GGM02c (Tapley et al. 2005) and the celestial mechanics approach, as proposed by Jäggi et al. (2007) and Prange et al. (2007) and applied to determine the satellite-only gravity field model AIUB-CHAMP01S.

2.2.2 Motivation for Regional Analysis

The new satellite missions with their innovative observation principles, as described in Section 2.1, in combination with the tailored analysis strategies, as introduced in Section 2.2.1, have provided an unprecedented increase in the accuracy of the gravity field solutions. So far, the most common way of representing the gravity field of the Earth has been an expansion in terms of spherical harmonics. Despite the outstanding results already achieved, it can be assumed that the signal content present in the satellite observations has not been fully exploited yet. The reasons for that are manifold, but one major aspect is the insufficient modeling of background forces (such as ocean tides or atmospheric variations), which is understood to be primarily responsible for the fact that, e.g., the projected GRACE baseline accuracy has not been achieved so far. Yet another reason for sub-optimal signal exploitation could be an insufficient modeling of the satellite data by a global representation by means of spherical harmonics. To extract the signal information present in the satellite and sensor data to full content, it seems reasonable to tailor the analysis process according to the specific characteristics of the gravity signal present in certain areas. Especially in the higher frequency part of the spectrum, the gravity field features vary significantly in different geographical regions. Those heterogeneities are caused by different topographic characteristics featuring rough gravity signal, for example caused by mountain areas or deep sea trenches, and rather smooth signal areas, for example in parts of the open oceans. In these cases, the heterogeneity of the gravity field cannot properly be taken into account with the help of spherical harmonics as basis functions with global support. Their resolution can only be defined globally, resulting in the problem that the maximum degree adequate for very rough gravity field features would cause instabilities in the computation procedure of the spherical harmonic coefficients. Therefore, it seems to be preferable to represent the global gravity field only up to a moderate, safely determinable degree and to model the additional detailed features by means of space localizing basis functions. The new analysis strategies described above, which take into account the in-situ type character of the observations, and which are tailored to densely observed short arcs of the satellite's orbit, have offered the possibility of using regionally restricted satellite data to explore the gravity field in the respective geographical areas. That is why, it seems to be reasonable to take advantage of this opportunity and to improve global solutions by regional refinement strategies. The approach proposed in this thesis is based on a global reference model represented by spherical harmonics and includes the calculation of regional refinements in certain geographical areas. These regional solutions should be modeled by means of radial basis functions that feature local (or nearly local) support. In this way, the resolution of the parameterization can be adapted according to the given gravity field signal in the region.

However, not only the parameterization itself can be adapted to the specific regional characteristics, but the complete analysis strategy should be adjusted to produce the best regionally adapted results. In this context, especially the regularization process necessary in connection with the downward continuation process has to be pointed out, as described, e.g., in EICKER et al. (2004), ILK et al. (2006), and EICKER et al. (2006). A global regularization characterized by a uniform global regularization parameter results in an overall filtering of the observations, which would lead to a mean dampening of the global gravity field features. Then the problem arises that in regions with rough gravity field characteristics there would still be additional information in the signal that is dampened unnecessarily strongly, whereas in regions with a smooth signal an even stronger regularization would be advisable. When dealing with space localizing basis functions, the regularization parameter can be chosen individually for different regions, as will be described in Chapter 5.2.5. Therefore, different regularization areas can be assigned, and their adapted regularization parameter can take into account the varying signal content. In this way, it is possible to extract more information from the given data than would be possible with global gravity field determination. Areas with a smooth gravity field signal, for example, can be regularized more strongly without dampening the signal, and in areas with strong high frequency signal no unnecessary dampening occurs. Results that support the significant improvements obtainable from such a regionally adapted regularization approach can be found in EICKER et al. (2007).

A further advantage of the regional refinement strategy is that it allows to combine the satellite data with additional data sets in a consistent way. The additional gravity field data (e.g. from altimeter observations, from airborne, or terrestrial data sets) can be included in the computation scheme in a straightforward step by using the satellite gravity solution as reference field and calculating regional refinements from the additional observations. Satellite gravity data usually covers the entire Earth, resulting in a uniformly distributed data coverage. But when surface data is to be taken into account as well, it has to be considered that the data could be sparse over some regions and very dense over other regions. Here an additional feature of the regional approach gets into focus, namely the advantage of dealing with regions with different data coverage more easily than possible in case of a global recovery strategy. If no data is available (e.g. when analyzing polar regions in case of the GOCE mission), the regional refinement may be skipped.

Furthermore, it has to be pointed out that the recovery of the gravity field does generally not represent a stand-alone purpose, but always serves the objective of being interpreted and applied afterwards by geoscientific users such as oceanographers, hydrologists, geophysicists etc. These related disciplines frequently deal with regionally limited phenomena and benefit from regional gravity field solutions exactly tailored to their area of interest. Besides the advantage mentioned above of enhanced signal extraction and tailored design of the radial basis functions, the regional solutions are easier to handle. Present efforts are being carried out to provide a high-resolution global gravity model, the EGM07 (Pavlis et al. 2005), which is to be expanded up to a spherical harmonic degree of 2160. To explore the full resolution of such a model, the user is faced with the computation effort of having to evaluate 4,669,921 parameters to be used to calculate every single point value. In case of space localizing basis functions, a comparable resolution can be obtained (for a limited area) with a significantly reduced number of parameters. Technically, it may be possible to calculate a global model of the given resolution when using tailored stable algorithms for the evaluation of the associated Legendre functions. But it can be assumed that spherical harmonic models cannot represent data of heterogeneous density and quality in a proper way, as pointed out, for example, in Schmidt et al. (2006).

If a spherical harmonic expansion is desired, regional solutions with global coverage can be merged to obtain a global solution whose spherical harmonic coefficients can be calculated, e.g. by means of quadrature methods. The maximum degree of this series expansion is only limited by the most detailed structures of the gravity field present in the regional solutions. This last step can be performed without stability problems and without losing the details of the regional gravity field features.

2.2.3 Existing Approaches to Regional Gravity Field Modeling

As emphasized in the last section, the calculation of regional gravity field solutions provides a wide variety of advantages when compared to the recovery of global models. Therefore, this opportunity is gaining more and more popularity, and can be considered as a very effective supplement to the conventional global models. While regional gravity field solutions have been routinely developed based on terrestrial (surface) data in the past, the application of regional modeling strategies have become meaningful only recently in the recovery process based on satellite data. In order to exploit the advantages of a regional analysis, the application of space localizing basis functions is inevitable. There is a considerably large variety of possibilities of choosing the proper basis functions to be used for the parameterization. In the following, some of the approaches in this context will be introduced. The description will thereby be restricted to applications that have been carried out based on the analysis of satellite data.

A rather early idea of modeling regional gravity fields from satellite data goes back to Weightmann (1965). It is based on the modeling of the gravitational potential of the Earth by superposition of the potential generated by point masses located in the interior of the Earth. The potential originating from a point mass depends on the reciprocal distance between the evaluation point and the respective mass element. Important applications of the idea are described, e.g., in Balmino (1972) or Barthelmes (1986).

Another approach is the representation of the gravity field by so-called mascons. The term 'mascons' stands for mass concentration blocks and refers to a small uniform layer of mass over a certain limited geographical region (such as a spherical block or cap). It was originally applied to specify the concentration of masses detected on the moon (Muller and Sjogren 1968). Concerning recent applications to the data of the satellite mission GRACE, the approach is, for example, explained in Rowlands et al. (2005). It is based on the representation of the difference in the gravitational potential caused by the individual mascons in terms of differential potential coefficients. The mascon parameter belonging to each of the regional areas is then a scaling factor to the set of potential coefficients associated with the mass element. To prevent the solution from numerical instabilities, spatial and temporal constraints are introduced to ensure pairs of mascons that are close to each other (either in time or in space) to stay close concering their values, too.

A widely-used idea in the modeling of potential fields are wavelets. A general description of the concept of wavelets can be found in literature, for instance in HOLSCHNEIDER (1995), FREEDEN et al. (1998), SCHMIDT (2001), or KELLER (2004). Wavelets represent the decomposition of a given signal into different frequency-dependent detail signals. Therefore, a wavelet representation is referred to as multi resolution representation. The basis functions consist of scaling functions and wavelet functions, each wavelet function representing the detail signal of a specific frequency band. They can be assumed to have local support in the space and in the time domain. The application of wavelets to the modeling of functions on the sphere leads to the application of spherical wavelets. The idea was applied to the analysis of CHAMP data by FENGLER et al. (2004) and SCHMIDT et al. (2005). SCHMIDT et al. (2006) used wavelet representations to model time variable gravity field information from GRACE data. FENGLER et al. (2007) used wavelets for the representation of time variable GRACE gravity fields. They made use of existing global spherical harmonic level 2 products and applied wavelets to the spherical harmonic coefficients. In this way the spherical harmonic solutions were decomposed into different scales.

An attempt to construct functions that concentrate a given signal as optimally as possible both in the time and the frequency domain is the so-called Slepian approach, see, e.g., SLEPIAN and POLLAK (1960) and LANDAU and POLLAK (1962). Since strict localization in the time and the frequency domain are mutually exclusive, the approach is based on the concept of maximizing the spatial concentration of a strictly band-limited function. This function can be found by maximizing the ratio of the norm of the bandlimited function within the concentration region and the norm of the respective function on the complete sphere; for a more detailed description of the derivation of the functions, refer to, for instance, SIMONS et al. (2006). In the context of analyzing the data of the new satellite missions, the concept of Slepian functions has been applied by BAUR and SNEEUW (2007) under special consideration of the polar gap problem occurring in case of the GOCE mission. Here the concentration region is a spherical band excluding the polar gap regions. Additional details of the use of Slepians in the context of the polar gap problem can be found in SIMONS and DAHLEN (2006). Furthermore, the Slepian functions have recently been applied to the study of regional phenomena from GRACE data (HAN and SIMONS 2008).

3. Modeling of the Regional Gravity Field

Within the course of the following chapter, the tools necessary for the modeling of regional gravity field solutions are introduced. In the first section, fundamental concepts of potential theory are summarized, as they present the basic setting in which the modeling has to take place. In the second section, the concept of reproducing kernel Hilbert spaces will be described. The introduction of spherical harmonics in the third section concludes the more fundamental part of the chapter. Essential for the modeling of regional gravity fields are space localizing basis functions; they will be described in the following section and specified as spherical splines. Smoothness characteristics implied by the design of the basis functions are investigated within the framework of the introduced reproducing kernel Hilbert spaces. The basis functions have to be arranged homogeneously on the sphere. Therefore, different possible choices of spherical point distributions are described in the last section, together with an investigation of their suitability as nodal point pattern for the spline functions.

3.1 Fundamentals of Potential Theory

In this section, important facts of the gravity field are summarized as far as they are relevant for this thesis. They establish the basic framework for the design of the basis functions that are used to represent the gravitational field. Further details on potential theory can be found, for example, in Heiskanen and Moritz (1967).

According to Newton's law of gravitation, one particle with the mass m_1 attracts a second particle with the mass m_2 with a force along the line of center of the two objects according to

$$\mathbf{K}_{12} = -Gm_1m_2\frac{\mathbf{r}_2-\mathbf{r}_1}{|\mathbf{r}_2-\mathbf{r}_1|^3}, \tag{3.1}$$

with the positions of the two masses $\mathbf{r}_1$ and $\mathbf{r}_2$ and the gravitational constant $G = (6672{\pm}4)10^{-14}m^3s^{-2}kg^{-1}$. The roles of attracting and attracted masses are interchangeable. Alternatively, the situation can be described by regarding the attracting mass m_Q to create a gravitational field strength around itself. This field strength can be evaluated at any given point $\mathbf{r}$ by

$$\mathbf{g}(\mathbf{r}) = -Gm_Q\frac{\mathbf{r}-\mathbf{r}_Q}{|\mathbf{r}-\mathbf{r}_Q|^3}. \tag{3.2}$$

An analog relation can be formulated for the other particle to be considered as source of gravitation. The force of gravity acting on a mass m within this force field can then be expressed by

$$\mathbf{K} = m\mathbf{g}. \tag{3.3}$$

As the gravitational forces of different point masses add up according to the superposition principle, the gravitational field strength of a solid body with the volume Σ can be given at a position $\mathbf{r}$ by

$$\mathbf{g}(\mathbf{r}) = -G\iiint\limits_{\Sigma}\rho(\mathbf{r}_Q)\frac{\mathbf{l}}{l^3}\,d\Sigma, \tag{3.4}$$

with the continuous density function $\rho(\mathbf{r}_Q)$ and the vector $\mathbf{l}$ with the length l between the field point $\mathbf{r}$ and the attracting particle $\mathbf{r}_Q$.

Gravity is a conservative force field; it can be derived as the gradient of a scalar potential, the gravitational potential $V(\mathbf{r})$, according to

$$\mathbf{g}(\mathbf{r}) = \nabla V(\mathbf{r}). \tag{3.5}$$

Consequently, the gravitational potential of a continuous body can be determined by

$$V(\mathbf{r}) = G \iiint\limits_{\Sigma} \rho(\mathbf{r}_Q) \frac{1}{l} \, d\Sigma, \tag{3.6}$$

where the integration constant is chosen such that the potential V vanishes at infinity. V is continuous throughout the whole space outside the masses and is differentiable arbitrarily often.

Of particular importance is the Laplace operator Δ, a differential operator that represents the sum of the second partial derivatives with respect to the three cartesian coordinates, as expressed by

$$\Delta V = \frac{\partial^2 V}{\partial x_1^2} + \frac{\partial^2 V}{\partial x_2^2} + \frac{\partial^2 V}{\partial x_3^2}. \tag{3.7}$$

When dealing with gravitation, two different cases are suitable to be considered separately according to the question whether a point is inside or outside of the attracting masses. It either satisfies Poisson's equation or Laplace's equation,

$$\Delta V = \begin{cases} -4\pi G\rho & \text{for } \mathbf{r} \text{ inside } \Sigma \quad \text{(Poisson)} \\ 0 & \text{for } \mathbf{r} \text{ outside } \Sigma \quad \text{(Laplace).} \end{cases} \tag{3.8}$$

Concerning the determination of the gravitational potential in the exterior of the Earth, Laplace's equation is relevant. Any function satisfying Laplace's equation at every point of a given region is called harmonic in this region. Outside of gravitational masses the gravitational potential is a harmonic function, whereas inside of those masses it is not harmonic, but satisfies Poisson's equation. In the following chapters of this thesis, only the harmonic case outside the masses will be of further interest. For the mathematical representation of the harmonic space, refer to Section 3.3.

3.2 Reproducing Kernel Hilbert Space

In Appendix A.1 the mathematical fundamentals dealing with function spaces are defined. Of particular interest is thereby the concept of a Hilbert space introduced as a function space and the space of square integrable functions $\mathcal{L}_2$. For specific applications in this thesis, the functions applied have to be restricted to those fulfilling specific smoothness requirements that are stronger than the one inflicted by Eq. (A.16) for functions belonging to $\mathcal{L}_2$. These smoothness conditions can be defined by a reproducing kernel, as will be described below. This is the reason why the respective spaces are denoted as reproducing kernel Hilbert spaces (RKHS). They are described, for example, in ARONSZAJN (1950) and MESCHKOWSKI (1962), and, for geodetic applications, in HEISKANEN and MORITZ (1967), KRARUP (1969), and TSCHERNING and RAPP (1974). The essential requirement for an RKHS is the fact that the evaluation functional F_x has to be bounded, i.e.

$$|F_x f| \leq M \, \|f\|, \tag{3.9}$$

with the concept of boundedness defined in equivalence with Eq. (A.24). An evaluation functional F_x assigns to a function f, belonging to the reproducing Hilbert space $\mathcal{H}_K$, its value at a certain point x according to

$$F_x f = f(x). \tag{3.10}$$

Eq. (3.9) implies that the value of an evaluation functional can be estimated by the norm of the function f. From this fact it can be concluded that in a general Hilbert space F_x cannot commonly be assumed to be bounded. For example, in the Hilbert space $\mathcal{L}_2$ outliers at single points change the value of $|F_x f|$, but do not change the norm of f. The function can be changed in a countable set of points without changing the norm, as single function values have no effect on the integral of the function.

The Riesz representation theorem (Riesz 1907), as given in Section A.2, applies to bounded linear functionals. The application to the bounded evaluation functional in an RKHS guarantees the existence of a function $K_x \in \mathcal{H}_K$ with

$$F_x f = f(x) = \langle K_x, f \rangle_K . \tag{3.11}$$

The symbol $\langle \cdot, \cdot \rangle_K$ denotes the inner product in $\mathcal{H}_K$, which will be specified below. According to the representation theorem, K_x is the unique representer of the evaluation functional F_x. The inner product of the representers of the evaluation functionals at two points x and y is given by

$$K(x,y) = \langle K_x, K_y \rangle_K , \tag{3.12}$$

with $K(x,y)$ being the reproducing kernel of the RKHS $\mathcal{H}_K$. The term 'reproducing kernel' is due to its reproducing property characterized by

$$f(y) = \langle K(x,y), f(x) \rangle_K . \tag{3.13}$$

Thus the inner product of the reproducing kernel with a function f reproduces the function itself. An RKHS is uniquely defined by the reproducing kernel (Theorem of Moore-Aronszajn, Aronszajn 1950).

In the following, it will be described how the reproducing kernel defines the smoothness condition imposed on functions in order to belong to $\mathcal{H}_K$. These considerations are closely connected to the definition of the inner product in the RKHS. In Appendix A.1 the inner product is defined in general, and a specification for the space $\mathcal{L}_2$ is given in Eq. (A.14) and Eq. (A.15). In the following, the $\mathcal{L}_2$ is considered on a bounded domain; in this case it is separable (concerning the concept of separability, refer to Appendix A.1). In $\mathcal{L}_2$ the scalar product of two functions is introduced, using Parseval's theorem, as the sum of the product of the Fourier coefficients of the two functions according to

$$\langle f, g \rangle_{\mathcal{L}_2} = \int f(x) g(x) dx = \sum_{i=0}^{\infty} f_i \cdot g_i . \tag{3.14}$$

The eigenvalue decomposition of linear operators in Appendix A.2.0.1 will be needed to understand the following concepts. According to Mercer's Theorem (Mercer 1909), every positive semi-definite kernel can be expressed by the orthonormal set of the eigenfunctions of the operator associated with the kernel. When the operator is interpreted as integral operator with kernel K, then the eigenvalue equation can be formulated as follows

$$\int K(x,y) u_i(y) dy = \lambda_i u_i(x), \tag{3.15}$$

where λ_i is the eigenvalue belonging to the respective eigenfunction u_i. In a separable Hilbert space, the kernel can be expressed by an expansion of these eigenfunctions,

$$K(x,y) = \sum_{i=0}^{\infty} \lambda_i \, u_i(x) \, u_i(y). \tag{3.16}$$

As mentioned in the context of the spectral decomposition of linear operators, the set of eigenfunctions is an orthogonal basis for the integral kernel spanning its range. Therefore, an arbitrary function within the range of the operator can be expanded with respect to the eigenfunctions in equivalence with Eq. (A.12). In the context of an RKHS, this implies that every function f belonging to $\mathcal{H}_K$ can also be expanded into a series of the eigenfunctions of K,

$$f = \sum_{i=0}^{\infty} f_i \, u_i . \tag{3.17}$$

Now the scalar product $\langle \cdot, \cdot \rangle_K$ shall be defined. In the framework of an RKHS, the product of the coefficients (3.14) is additionally divided by the respective eigenvalue of the reproducing kernel according to

$$\langle f, g \rangle_K = \left\langle \sum_{i=0}^{\infty} f_i u_i, \sum_{i=0}^{\infty} g_i u_i \right\rangle_K = \sum_{i=0}^{\infty} \frac{f_i \cdot g_i}{\lambda_i} . \tag{3.18}$$

Then the definition of the norm of a function f is given by

$$\|f\|_K^2 = \langle f, f\rangle_K = \sum_{i=0}^{\infty} \frac{f_i^2}{\lambda_i}. \tag{3.19}$$

The RKHS can be considered as the closure of the set of all functions fulfilling the condition $\|\cdot\|_K < \infty$. Therefore, for f to be in the RKHS defined by K, Eq. (3.19) has to be finite. This imposes the smoothness condition mentioned above on the function, as the coefficients of the function f have to approach zero sufficiently fast to guarantee

$$\sum_{i=0}^{\infty} \frac{f_i^2}{\lambda_i} \qquad < \infty. \tag{3.20}$$

This implies that the coefficients f_i have to descend faster than the coefficients λ_i of the kernel. Thus the functions as elements of an RKHS have to be smoother than the kernel itself. This results in stronger limitations being imposed on the smoothness of the function f, and thus on the decreasing behavior of the coefficients, than it is the case in $\mathcal{L}_2$, as can be concluded from the comparison of Eq. (3.20) with Eq. (A.16). In the following, the scalar product denoted by $\langle\cdot,\cdot\rangle$ and the norm denoted by $\|\cdot\|$ will refer to the respective calculations with respect to $\mathcal{L}_2$ without the use of the subscript.

In finite dimensional spaces, the kernel K becomes the finite dimensional matrix $\mathbf{K}$. The division by the singular values in the scalar product given by Eq. (3.18) corresponds to the introduction of the inverse matrix $\mathbf{K}^{-1}$. The scalar product of the two $n \times 1$ vectors $\mathbf{f} = (f_1, ..., f_n)^T$ and $\mathbf{g} = (g_1, ..., g_n)^T$ is then defined by

$$\langle \mathbf{f}, \mathbf{g}\rangle_K = \left\langle \mathbf{f}, \mathbf{K}^{-1}\mathbf{g}\right\rangle. \tag{3.21}$$

Using the definition of the scalar product, as given by Eq. (3.18), demonstrates the reproducing character of the kernel

$$\langle f(x), K(x,y)\rangle_K = \left\langle \sum_{i=0}^{\infty} f_i u_i(x), \sum_{i=0}^{\infty} \lambda_i u_i(x) u_i(y) \right\rangle_K = \sum_{i=0}^{\infty} \frac{1}{\lambda_i} f_i \ \lambda_i \ u_i(y) = \sum_{i=0}^{\infty} f_i \ u_i(y) = f(y). \tag{3.22}$$

To provide a better understanding of the concept of reproducing kernel Hilbert spaces, again the difference between an RKHS and $\mathcal{L}_2$ shall be demonstrated. In $\mathcal{L}_2$ the analogy to a reproducing kernel is the Dirac delta functional, its reproducing property can be stated symbolically by

$$f(y) = \langle f(x), \delta(x,y)\rangle = \int f(x)\delta(x,y)dx. \tag{3.23}$$

According to this equation, the Dirac functional is the evaluation functional for the scalar product defined in $\mathcal{L}_2$. But it is not an element of $\mathcal{L}_2$, and it is not bounded. Therefore, $\mathcal{L}_2$ is not an RKHS. The Hilbert space $\mathcal{L}_2$ contains too many functions that are not smooth enough. The RKHS consists only of those functions satisfying the smoothness condition given by Eq. (3.20), which is a stronger requirement than the one imposed by Eq. (A.16). Thus the RKHS implied by a specific kernel K is smaller than $\mathcal{L}_2$, with different kernels introducing different characteristics of smoothness. Every positive definite function is a reproducing kernel for some RKHS, thus for any given positive definite kernel an RKHS can be constructed.

3.3 Spherical Harmonics

In the following, spherical harmonics as frequently-used tool for the representation of global data on the sphere are introduced. The use of spherical harmonics in gravity field recovery for the purpose of representing global fields is a well-established technique. Consequently, most of the existing gravity field models are formulated in terms of spherical harmonic coefficients. Spherical harmonics are the eigenfunctions of the Laplace operator given in Eq. (3.7) and represent a set of basis functions with global support. Therefore, they are well-suited for the representation of global gravity field solutions. Their applicability to the modeling of regional fields has to be regarded as less appropriate. Within the context of regional gravity field recovery as treated in this thesis, the understanding of spherical harmonics is nonetheless inevitable. On the one hand, it helps to point out the differences when dealing with space localizing basis functions. On the other hand, the construction of the basis functions, as will be described in Section 3.4, requires concepts applied in connection with spherical harmonics, too. In the first part of this section, a mathematical description of spherical harmonics will be given, introducing them as the restriction of harmonic polynomials to the unit sphere, as described, e.g., by FREEDEN et al. (1998). Subsequently, the spherical harmonics will be adapted for geodetic applications, refer to HEISKANEN and MORITZ (1967). A further comprehensive treatment of the subject of spherical harmonics can be found in HOBSON (1931). As an introduction to the concept of space localizing basis functions, the covariance function on the sphere is then described in the third part of this section, followed by the application of the concept of reproducing kernel Hilbert spaces to the sphere.

3.3.1 Mathematical Description

In the following, vectors $\mathbf{x}$ and $\mathbf{y}$ will be used to denote vectors on the unit sphere $|\mathbf{x}| = |\mathbf{y}| = 1$. This will prove to be helpful, as most of the following considerations will be performed with respect to the unit sphere Ω with

$$\Omega = \{\mathbf{x} \in \mathbb{R}^3 \mid |\mathbf{x}| = 1\}. \tag{3.24}$$

A vector $\mathbf{x}_r \in \mathbb{R}^3, \mathbf{x}_r = (rx_1, rx_2, rx_3)^T$ can then be represented by $\mathbf{x}_r = r\mathbf{x}$ with $|\mathbf{x}_r| = r$.

A polynomial H_n is homogeneous of degree n in x_1, x_2, x_3 if

$$H_n(r\mathbf{x}) = H_n(rx_1, rx_2, rx_3) = r^n H_n(x_1, x_2, x_3). \tag{3.25}$$

A homogeneous polynomial H_n is harmonic if it fulfills Laplace's equation, as defined in Eq. (3.7),

$$\Delta H_n = 0. \tag{3.26}$$

The restriction of any homogeneous harmonic polynomial H_n of degree n to the unit sphere Ω is called a (Laplace's) surface spherical harmonic Y_n of degree n,

$$Y_n = H_n|_\Omega. \tag{3.27}$$

It can be proven that if $r^n Y_n(\mathbf{x})$ is harmonic, then $\frac{1}{r^{n+1}} Y_n(\mathbf{x})$ is harmonic as well. These functions

$$\frac{1}{r^{n+1}} Y_n(\mathbf{x}) \tag{3.28}$$

are called solid spherical harmonics of degree n. They represent the solution to Eq. (3.26), cf. HEISKANEN and MORITZ (1967). On the unit sphere the following inner product can be defined,

$$\langle u(\mathbf{x}), v(\mathbf{x}) \rangle = \frac{1}{4\pi} \iint_\Omega u(\mathbf{x})\ v(\mathbf{x})\ d\Omega. \tag{3.29}$$

A basic property of spherical harmonics is orthogonality. If $Y_n(\mathbf{x})$ and $Y_{\bar{n}}(\mathbf{x})$ are spherical harmonics of degree n and $\bar{n}$, respectively, they are orthogonal in the sense of the inner product defined by (3.29),

$$\langle Y_n, Y_{\bar{n}} \rangle = \frac{1}{4\pi} \iint\limits_{\Omega} Y_n(\mathbf{x}) Y_{\bar{n}}(\mathbf{x})\, d\Omega = 0, \ n \neq \bar{n}. \tag{3.30}$$

For every Y_n there exist $2n+1$ linearly independent orthogonal basis functions denoted by Y_{nm} with $m \in [-n, n]$. Consequently, the space of spherical harmonics of degree n is of the dimension $\dim Y_n = 2n+1$, and Y_n can be separated into a series expansion of these basis functions according to

$$Y_n(\mathbf{x}) = \sum_{m=-n}^{n} c_{nm} Y_{nm}(\mathbf{x}). \tag{3.31}$$

The basis functions Y_{nm} are chosen to be orthonormal with respect to the inner product (3.29),

$$\langle Y_{nm}(\mathbf{x}), Y_{\bar{n}\bar{m}}(\mathbf{x}) \rangle = \frac{1}{4\pi} \iint\limits_{\Omega} Y_{nm}(\mathbf{x}) Y_{\bar{n}\bar{m}}(\mathbf{x})\, d\Omega = \delta_{n\bar{n}} \delta_{m\bar{m}}. \tag{3.32}$$

Consequently, the coefficients c_{nm} can be calculated by

$$c_{nm} = \langle Y_n(\mathbf{x}), Y_{nm}(\mathbf{x}) \rangle = \frac{1}{4\pi} \iint\limits_{\Omega} Y_n(\mathbf{x}) Y_{nm}(\mathbf{x})\, d\Omega. \tag{3.33}$$

The basis functions Y_{nm} are also called surface spherical harmonics of degree n and order m. It has to be distinguished, depending on the specific context, whether the term 'spherical harmonic' is referred to (Laplace's) spherical harmonics Y_n or their basis functions Y_{nm}.

Of considerable interest is the addition theorem of spherical harmonics,

$$\sum_{m=-n}^{n} Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{y}) = \sqrt{2n+1}\, P_n(\mathbf{x} \cdot \mathbf{y}), \tag{3.34}$$

which establishes a relationship between the (fully normalized) Legendre polynomials $P_n(\mathbf{x} \cdot \mathbf{y}) = P_n(\cos(\psi))$ defined on the interval $[-1, 1]$ on the one hand and the functions $Y_{nm}(\mathbf{x})$ defined on the sphere Ω on the other hand. The spherical distance ψ between two points is related to the inner product of the two points by

$$t = \cos\psi = \mathbf{x} \cdot \mathbf{y}, \tag{3.35}$$

and the Legendre polynomials $P_n(t)$ only depend on this spherical distance. Their fully normalized version is given by

$$\frac{1}{2} \int_{-1}^{1} P_n(t) P_{\bar{n}}(t)\, dt = \delta_{n\bar{n}}. \tag{3.36}$$

A connection between the Legendre Polynomials $P_n(\mathbf{x} \cdot \mathbf{y})$ and the Laplace's spherical harmonics Y_n can be established when both sides of Eq. (3.34) are multiplied by $Y_n(\mathbf{x})$ and subsequently an integration over the unit sphere is performed. This leads to the following relation,

$$Y_n(\mathbf{y}) = \frac{\sqrt{2n+1}}{4\pi} \iint\limits_{\Omega} Y_n(\mathbf{x}) P_n(\mathbf{x} \cdot \mathbf{y}) d\Omega. \tag{3.37}$$

Any function $f(\mathbf{x}) \in \mathcal{L}_2(\Omega)$ on Ω can be developed into a series of spherical harmonics. In this context, the spherical harmonics serve as orthogonal basis functions (as given by Eq. (A.12) for arbitrary Hilbert spaces),

$$f(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm} Y_{nm}(\mathbf{x}). \tag{3.38}$$

Eq. (3.38) is denoted as spherical harmonic synthesis. When the spherical harmonic coefficients are to be calculated from a given function by using the orthogonality property of surface spherical harmonics,

$$c_{nm} = \frac{1}{4\pi} \iint\limits_{\Omega} f(\mathbf{x}) Y_{nm}(\mathbf{x})\, d\Omega, \tag{3.39}$$

this is known as spherical harmonic analysis. A harmonic function $f(\mathbf{x}_r)$ in the exterior of Ω can be expanded in terms of solid spherical harmonics by

$$f(\mathbf{x}_r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} \sum_{m=-n}^{n} c_{nm} Y_{nm}(\mathbf{x}). \tag{3.40}$$

According to Parseval's theorem, the norm of the function f (in the $\mathcal{L}_2$ sense) can be phrased in terms of the spherical harmonic coefficients by

$$\|f(\mathbf{x})\|^2 = \frac{1}{4\pi} \iint\limits_{\Omega} f(\mathbf{x})^2 d\Omega = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=0}^{\infty} \sigma_n^2. \tag{3.41}$$

Here σ_n^2 are called degree variances of degree n. The expansion of a function in spherical harmonics according to (3.38) can be interpreted as Fourier expansion on the sphere, with each degree and order being associated with a certain frequency.

3.3.2 The Use of Spherical Harmonics in Geodesy

The above specifications of spherical harmonics have taken place from a merely mathematical point of view. In the following, the spherical harmonic expansion will be adapted to geodetic problems such as the task of describing functionals of the gravitational potential of the Earth, as introduced, for example, by HEISKANEN and MORITZ (1967). While the mathematical considerations above have been performed with respect to the unit sphere, a series expansion of the gravitational potential of the Earth is commonly expressed in the exterior of a sphere Ω_R with a radius R being the mean radius of the equator of the Earth. The potential in the exterior of this sphere can then be represented in terms of solid spherical harmonics by

$$V(\mathbf{x}_r) = \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} c_{nm} Y_{nm}(\mathbf{x}), \tag{3.42}$$

with G denoting the gravitational constant and M representing the mass of the Earth. The factor GM/R is introduced, so that the coefficients c_{nm} can be identified as dimensionless values, and the potential $V(\mathbf{x}_r)$ is provided with the correct units.

So far, the fully normalized spherical harmonics have been used with the orthonormality relation given by Eq. (3.32). Quite common in geodetic literature is the use of the not normalized version, here denoted by $\hat{Y}_{nm}$. It is related to the fully normalized surface spherical harmonics Y_{nm} by

$$Y_{nm}(\mathbf{x}) = \sqrt{(2-\delta_{0m})(2n+1)\frac{(n-|m|)!}{(n+|m|)!}}\hat{Y}_{nm}(\mathbf{x}). \tag{3.43}$$

In the case of the Earth's gravity field, the norm of a function as introduced in Eq. (3.41) can be approximated by an empirical model known as Kaula's rule of thumb (KAULA 1966),

$$\|V(\mathbf{x})\|^2 = \sum_{n=0}^{\infty} \sigma_n^2 \approx \sum_{n=0}^{\infty} (2n+1)\frac{10^{-10}}{n^4}. \tag{3.44}$$

Here $\frac{10^{-10}}{n^4}$ corresponds to one single coefficient of degree n, and the factor $(2n+1)$ takes care of the fact that there are $(2n+1)$ coefficients per degree. Then the standard deviations per degree can be approximated by

$$\sigma_n = \sqrt{(2n+1)\frac{10^{-10}}{n^4}}. \tag{3.45}$$

In geodetic applications, the introduction of spherical coordinates, radius r, polar distance or co-latitude ϑ, and geocentric longitude λ is quite common. It holds for the relationship to the rectangular coordinates

$$\mathbf{x}_r = (r\cos\lambda\sin\vartheta, r\sin\lambda\sin\vartheta, r\cos\vartheta)^T. \tag{3.46}$$

The basis functions are then chosen according to

$$Y_{nm}(\mathbf{x}) = \left\{ \begin{array}{lllll} C_{nm}(\mathbf{x}) & = P_{nm}(\cos\vartheta)\cos(m\lambda) & \text{for} & m = 0,...,n \\ S_{nm}(\mathbf{x}) & = P_{n|m|}(\cos\vartheta)\sin(|m|\,\lambda) & \text{for} & m = -n,...,-1 \end{array} \right. , \tag{3.47}$$

with the associated Legendre functions $P_{n|m|}(\cos\vartheta)$. The coefficients assigned to the $C_{nm}(\mathbf{x})$ and $S_{nm}(\mathbf{x})$ are denoted by c_{nm} and s_{nm}, respectively. This results in the following expression for the gravitational potential,

$$V(\mathbf{x}_r) = \frac{GM}{R}\sum_{n=0}^{\infty}\left(\frac{R}{r}\right)^{n+1}\sum_{m=0}^{n} c_{nm}C_{nm}(\mathbf{x}) + s_{nm}S_{nm}(\mathbf{x}). \tag{3.48}$$

It has to be pointed out that the c_{nm} in (3.48) are different from those applied in (3.38). In most of the following investigations within this thesis, the expansion of a function in terms of spherical harmonics according to Eq. (3.38) will be used. However, when dealing with the discrete orthogonality relations of spherical harmonics that will be exploited in Section 6.2.1, the differentiation between sine and cosine functions is significant. Therefore, at that point, the use of Eq. (3.48) will be more reasonable.

3.3.3 Covariance Function on the Sphere

The covariance function characterizes the statistical properties of the gravity field. It can be applied in signal modeling or accuracy estimation. In the context of this thesis, it is of particular interest for the design of the radial basis functions in Section 3.4, as they are directly based on the statistical characteristics given by the covariance function. When defining the covariance function, the gravity field functionals have to be interpreted as signals of a stationary stochastic process on the sphere. Concerning the discussion whether this is a valid assumption, see, for example, Moritz (1980) and Moritz and Sansò (1981). In case of the gravity field, there is only one realization of the stochastic process available. Therefore, it has to be assumed to be ergodic, which implies that the average in time can be taken to be equal to the average over the ensemble, in this case the average over the sphere. This allows for the statistical characteristics to be derived from one single realization. Furthermore, homogeneity and isotropy are presumed resulting in a dependency only on the spherical distance between two points. The expectation value of the homogeneous, isotropic, and ergodic stochastic process on the sphere is given by the average (Moritz 1978),

$$M(f) = \frac{1}{4\pi}\int_{\lambda=0}^{2\pi}\int_{\vartheta=0}^{\pi} f(\vartheta,\lambda)\sin(\vartheta)d\vartheta d\lambda, \tag{3.49}$$

with $M(f)$ denoting the average operator applied to the function f. The average (3.49) becomes zero if the corresponding expansion of the function in a series of spherical harmonics does not include a term of degree $n = 0$. This will be presumed in later applications throughout this thesis by introducing a reference field whose mass is set equal to the mass of the Earth. Furthermore, the harmonic terms of degree one are assumed to become zero by an appropriate choice of the origin of the coordinate system. Therefore, in the expansion of the covariance functions below, the zeroth and first order spherical harmonics will be omitted.

The covariance function $\mathcal{C}(\mathbf{x},\mathbf{y})$ can be calculated by averaging the product of two function values according to

$$\mathcal{C}(\mathbf{x},\mathbf{y}) = \mathcal{C}(\psi) = M(f(\mathbf{x}), f(\mathbf{y})) = \frac{1}{8\pi^2} \int\limits_{\lambda=0}^{2\pi} \int\limits_{\vartheta=0}^{\pi} \int\limits_{\alpha=0}^{2\pi} f(\vartheta,\lambda) f(\vartheta',\lambda') \sin(\vartheta) d\vartheta d\lambda d\alpha, \tag{3.50}$$

and it only depends on the spherical distance ψ between these two points. Here (ϑ,λ) are the spherical coordinates of point $\mathbf{x}$, and (ϑ',λ') are the spherical coordinates of point $\mathbf{y}$. The (ϑ',λ') depend on (ϑ,λ) due to the fact that the two points on the sphere are separated by the spherical distance ψ and that $\mathbf{y}$ is located at azimuth α from $\mathbf{x}$. The integration over λ and ϑ implies homogeneity, and the integration over α denotes isotropy. The averaging operator $M(f(\mathbf{x}), f(\mathbf{y}))$ can be interpreted as averaging first over a circle with radius ψ located at $\mathbf{x}$ and then varying $\mathbf{x}$ over the whole sphere. Following the derivations described by HEISKANEN and MORITZ (1967), the covariance function can be expanded into a series of spherical harmonics. Since it depends only on the spherical distance between two points, it can be expressed by Legendre polynomials,

$$\mathcal{C}(\mathbf{x},\mathbf{y}) = \mathcal{C}(\psi) = \sum_{n=2}^{\infty} c_n P_n(\cos\psi). \tag{3.51}$$

The coefficients c_n can be obtained by

$$c_n = \frac{1}{2} \int_{\psi=0}^{\pi} \mathcal{C}(\psi) P_n(\cos\psi) \sin\psi \, d\psi. \tag{3.52}$$

Inserting Eq. (3.50) into Eq. (3.52) leads to

$$c_n = \frac{1}{4\pi}\frac{1}{4\pi} \int_{\lambda=0}^{2\pi} \int_{\vartheta=0}^{\pi} \int_{\alpha=0}^{2\pi} \int_{\psi=0}^{\pi} f(\vartheta,\lambda) f(\vartheta',\lambda') P_n(\cos\psi) \sin\psi \, d\psi \, d\alpha \cdot \sin\vartheta \, d\vartheta \, d\lambda. \tag{3.53}$$

At first the integration with respect to α and ψ shall be considered. Taking into account the addition theorem of Eq. (3.34) and the spherical harmonic analysis defined in Eq. (3.39) and introducing a change of the integration variables, the integration can be carried out as follows,

$$\begin{aligned} &\frac{1}{4\pi} \int_{\alpha=0}^{2\pi} \int_{\psi=0}^{\pi} f(\vartheta',\lambda') P_n(\cos\psi) \sin\psi \, d\psi d\alpha \\ &\qquad = \frac{1}{4\pi} \int_{\lambda'=0}^{2\pi} \int_{\vartheta'=0}^{\pi} f(\vartheta',\lambda') P_n(\cos\psi) \sin\vartheta' \, d\psi d\lambda' = \frac{f_n(\vartheta,\lambda)}{\sqrt{2n+1}}. \end{aligned} \tag{3.54}$$

This leads to the formulation of the coefficients c_n by

$$c_n = \frac{1}{4\pi} \frac{1}{\sqrt{2n+1}} \int_{\lambda=0}^{2\pi} \int_{\vartheta=0}^{\pi} f(\vartheta,\lambda) f_n(\vartheta,\lambda) \sin\vartheta \, d\vartheta \, d\lambda. \tag{3.55}$$

Here f_n denotes the Laplace's surface spherical harmonics given by Eq. (3.31), written in the form

$$f(\vartheta,\lambda) = \sum_{n'=2}^{\infty} f_{n'}(\vartheta,\lambda). \tag{3.56}$$

When inserting Eq. (3.56) into Eq. (3.55), it follows by applying the orthogonality relations defined by Eq. (3.30),

$$c_n = \frac{1}{\sqrt{2n+1}} \sum_{m=-n}^{n} c_{nm}^2 = \frac{\sigma_n^2}{\sqrt{2n+1}}. \tag{3.57}$$

Thus the coefficients of the covariance function are the variances of the Laplace spherical harmonics of degree n, the so-called degree variances as introduced in Eq. (3.41). Then the covariance function is given by

$$\mathcal{C}(\mathbf{x},\mathbf{y}) = \sum_{n=2}^{\infty} \frac{\sigma_n^2}{\sqrt{2n+1}} P_n(\mathbf{x}\cdot\mathbf{y}) = \sum_{n=2}^{\infty} \frac{\sigma_n^2}{2n+1} \sum_{m=-n}^{n} Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{y}). \tag{3.58}$$

The degree variances represent the power spectral density of the function f, thus they give evidence of how much energy is contained in a certain frequency n.

3.3.4 RKHS on the Sphere

The concept of a Hilbert space with reproducing kernel, introduced in Section 3.2, can be applied to functions on the sphere, too. In geophysical applications, the functions to be determined are typically bounded and continuous, which implies that they cannot be arbitrarily rough. Therefore, the space $\mathcal{L}_2(\Omega)$ of square integrable functions on the sphere contains too many functions that are not smooth enough, as square integrable functions are not necessarily continuous or bounded. Because of this reason, the restriction of the functions in question to those satisfying certain smoothness conditions imposed by a reproducing kernel becomes necessary for certain applications. A reproducing kernel on the unit sphere can be expressed as a series of Legendre polynomials according to

$$K(\mathbf{x},\mathbf{y}) = \sum_{n=0}^{\infty} \sqrt{2n+1}\,\lambda_n\, P_n(\mathbf{x}\cdot\mathbf{y}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \lambda_n Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{y}), \tag{3.59}$$

with the spherical harmonics serving as orthogonal eigenfunctions as introduced by Eq. (3.16). For a function f to be element of the RKHS, the norm as defined by (3.19) has to be finite. This imposes the smoothness condition on the function, as its coefficients have to descend sufficiently fast to ensure

$$\|f(\mathbf{x})\|_K^2 = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{c_{nm}^2}{\lambda_n} < \infty. \tag{3.60}$$

As already discussed in Section 3.2, the term 'smoothness' of a function refers to the decreasing behavior of the coefficients. The faster the coefficients c_{nm} descend with increasing n, the less energy is contained in the higher frequencies. The inner product corresponding to (3.60) is then defined for two functions $f(\mathbf{x})$ and $g(\mathbf{x})$ according to Eq. (3.18) by

$$\langle f(\mathbf{x}), g(\mathbf{x})\rangle_K = \left\langle \sum_{n=0}^{\infty}\sum_{m=-n}^{n} f_{nm} Y_{nm}(\mathbf{x}), \sum_{n=0}^{\infty}\sum_{m=-n}^{n} g_{nm} Y_{nm}(\mathbf{x}) \right\rangle_K = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{g_{nm} f_{nm}}{\lambda_n}. \tag{3.61}$$

Referring to this inner product, the reproduction property of the kernel leads to

$$\langle K(\mathbf{x},\mathbf{y}), f(\mathbf{x})\rangle_K = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{1}{\lambda_n}\,\lambda_n\, f_{nm} Y_{nm}(\mathbf{y}) = f(\mathbf{y}). \tag{3.62}$$

Generally, in mathematical considerations, the smoothness of a function is defined by its differentiability, meaning the more often a function can be differentiated, the smoother it is. To illustrate the relationship between the smoothness requirements imposed by a reproducing kernel and the notion of differentiability, the first radial derivative of a function expanded into a series of solid spherical harmonics according to (3.40) shall be investigated on the unit sphere,

$$\left.\frac{df(\mathbf{x}_r)}{dr}\right|_{r=1} = \sum_{n=0}^{\infty} -(n+1) \sum_{m=-n}^{n} c_{nm} Y_{nm}(\mathbf{x}). \tag{3.63}$$

The amplification of the higher frequencies due to the introduction of the factor $(n+1)$ becomes evident. A function f has a derivative in $\mathcal{L}_2$ if the series converges

$$\left\|\frac{df(\mathbf{x}_r)}{dr}\right\|^2 = \sum_{n=0}^{\infty}(n+1)^2 \sum_{m=-n}^{n} c_{nm}^2 \qquad < \infty. \tag{3.64}$$

Comparing (3.60) and (3.64) reveals that the demand of differentiability in this case is equivalent to requiring that the function f belongs to an RKHS with the eigenvalues $\lambda_n = 1/(n+1)$ of the reproducing kernel. Thus it can be concluded that the smoothness requirements induced by the demand of a function to belong to a certain RKHS and to have a corresponding differentiability are directly related concepts. For the function to have a first derivative in $\mathcal{L}_2$, the degree variances $\sum_{m=-n}^{n} c_{nm}^2$ have to decrease sufficiently fast with increasing degree n to ensure convergence despite division by the descending eigenvalues. Higher order derivatives are represented equivalently by kernels with eigenvalues that descend even faster. However, it has to be pointed out that these considerations are valid for functions on the unit sphere. In the outer space, harmonic functions are differentiable infinitely often and the norm of the function as well as the norm of all its derivatives have to converge.

Regarding (3.58), under consideration of (3.57), it can be observed that the covariance function gives evidence of the smoothness of the associated function. This leads to the idea of using the covariance function $\mathcal{C}(\mathbf{x},\mathbf{y})$ of (3.58) itself as a reproducing kernel to define an RKHS $\mathcal{H}_\mathcal{C}$ with the corresponding smoothness characteristics (Tscherning 1972). In this case, the eigenvalues to be inserted into (3.59) have to be chosen according to

$$\lambda_n = \frac{\sigma_n^2}{2n+1}. \tag{3.65}$$

As for the covariance function the coefficients of degree zero and degree one have been assumed to be zero, this has to be taken into account in the following considerations as well. Since a reproducing kernel has to be positive definite, its eigenvalues have to be positive. Therefore, in this case the series expansion of the kernel has to start with degree $n = 2$ to ensure positive definiteness. This implies that the corresponding eigenfunctions Y_{nm} span a Hilbert space of functions for which the zeroth and the first moment vanish,

$$\iint\limits_\Omega f(\mathbf{x})\, d\Omega = 0 \qquad \text{and} \qquad \iint\limits_\Omega x_\alpha f(\mathbf{x})\, d\Omega = 0 \qquad \text{with} \qquad x_\alpha = \{x, y, z\}\,. \tag{3.66}$$

In equivalence with Eq. (3.62), the reproducing property can be specified for this kernel as well,

$$\langle \mathcal{C}(\mathbf{x},\mathbf{y}), f(\mathbf{x})\rangle_\mathcal{C} = \sum_{n=2}^{\infty}\sum_{m=-n}^{n} \frac{2n+1}{\sigma_n^2}\frac{\sigma_n^2}{2n+1} f_{nm}Y_{nm}(\mathbf{y}) = f(\mathbf{y}). \tag{3.67}$$

This is valid for an arbitrary function $f(\mathbf{x}) = \sum_{n=2}^{\infty}\sum_{m=-n}^{n} f_{nm}Y_{nm}(\mathbf{x})$, as long as it is an element of $\mathcal{H}_\mathcal{C}$. If, on the contrary, the norm $\|\cdot\|_\mathcal{C}$ is calculated for the function (e.g. the gravitational potential) from which the covariance function and thus the coefficients (3.65) are derived, this yields

$$\|V(\mathbf{x})\|_\mathcal{C}^2 = \sum_{n=2}^{\infty}\frac{2n+1}{\sigma_n^2}\sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=2}^{\infty}\frac{2n+1}{\sigma_n^2}\cdot\sigma_n^2 = \sum_{n=2}^{\infty}(2n+1) = \infty. \tag{3.68}$$

Obviously, this contradicts the smoothness condition imposed by the kernel. Thus it can be concluded that the function, as for example the gravitational potential of the Earth, does not belong to the RKHS defined by its covariance function. This problem as a consequence of choosing the covariance function as reproducing kernel was also pointed out by Tscherning (1977). A way out of this problem would be to slightly modify the degree variances by substituting σ_n^2 by $\sigma_n^2(1+\epsilon)^n$ with a small positive quantity ϵ, as proposed in Moritz (1980). This results in the modified (and thus slightly rougher) kernel function

$$\bar{\mathcal{C}}(\mathbf{x},\mathbf{y}) = \sum_{n=2}^{\infty}\frac{\sigma_n^2(1+\epsilon)^n}{2n+1}\left(\frac{R}{R'}\right)^n\sum_{m=-n}^{n} Y_{nm}(\mathbf{x})Y_{nm}(\mathbf{y}). \tag{3.69}$$

The parameter ϵ can adjust the smoothness of the reproducing kernel. The consideration of the covariance function as reproducing kernel will be discussed in the context of regularization in Section 5.2. In Section 7.2 the effect of a modified covariance function will be investigated. Despite the fact that the problem of convergence is of theoretical interest, it does not seem to have a significant impact on the results of practical applications.

3.3.5 Upward Continuation

The above considerations refer to the unit sphere. Introducing solid spherical harmonics, functions can also be formulated in the exterior of a sphere with arbitrary radius R, as indicated in Eq. (3.42). This upward continuation from the sphere to a point $\mathbf{x}_r$ in its exterior with a distance r from the center of the sphere arises as the solution of a first boundary value problem of potential theory, see, e.g., Heiskanen and Moritz (1967). It can be formulated as the application of an upward continuation operator. The kernel of this operator can be given by

$$U(\mathbf{x}, \mathbf{x}_r) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{R} \left(\frac{R}{r}\right)^{n+1} Y_{nm}\left(\frac{\mathbf{x}_r}{r}\right) Y_{nm}(\mathbf{x}). \tag{3.70}$$

Then the function in the exterior of the sphere can be expressed by applying the upward continuation operator to the respective function on the sphere,

$$f(\mathbf{x}_r) = \iint_{\Omega} U(\mathbf{x}, \mathbf{x}_r) f(\mathbf{x}) d\Omega = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{R} \left(\frac{R}{r}\right)^{n+1} c_{nm} Y_{nm}\left(\frac{\mathbf{x}_r}{r}\right). \tag{3.71}$$

The formulation of the upward continuation as operator equation will prove to be helpful when investigating its spectral properties, especially with respect to the ill-posedness of the downward continuation process that will be dealt with in Section 5.1.1.

3.4 Space Localizing Basis Functions

The following section will introduce the concept of space localizing basis functions, as they are an essential tool in regional gravity field modeling. At first, some remarks are to be made concerning the subject of localization, because this helps to point out the differences of space localizing basis functions compared to the spherical harmonics that were described in Section 3.3 as basis functions especially for global gravity field modeling.

A function can be localizing in the space as well as in the frequency domain. The term 'localization' refers to the size of the area in the specific domain in which the function does not vanish. The smaller this area is, the better the function localizes. According to Heisenberg's uncertainty principle (see, for example, Heitz and Stöcker-Meier 1994), a function cannot have perfect localizing properties in the frequency domain and at the same time in the space domain. The ratio of both localizations is reciprocally proportional, as the better the function localizes in one domain, the worse does it localize in the other domain. A schematic illustration of this matter is given in Fig. 3.1; here also the different basis functions are assigned to their respective localization properties. The spherical harmonics described above, for instance, have perfect localizing properties in the frequency domain, as every degree n and order m can be associated with one single frequency. Consequently, they do not have any localizing features in the space domain at all. Spherical harmonics are basis functions with global support, resulting in the fact that each spherical harmonic is significantly different from zero almost everywhere on the sphere. Therefore, changes in one coefficient always affect the complete sphere. On the other hand, changes in a regionally restricted area of the represented field will affect the whole set of spherical harmonic coefficients. Consequently, the calculation of one single field value requires the evaluation of every coefficient of the global model. The directly opposite characteristics provide the Dirac functionals. They are different from zero in only one single point in the space domain, thus providing perfect localizing properties there. On the other hand, they contain every frequency with equal weight and, therefore, exhibit no localization in the frequency domain at all. A compromise between space and frequency localization is provided by kernels such as those that will be applied as basis functions for the regional modeling in this thesis. Despite no perfect localization in the space domain, they are different from zero only in a geographically limited area or at least decay fast enough in the space domain to be regarded as quasi space localizing. Here it can be distinguished between non-bandlimited kernels and bandlimited ones

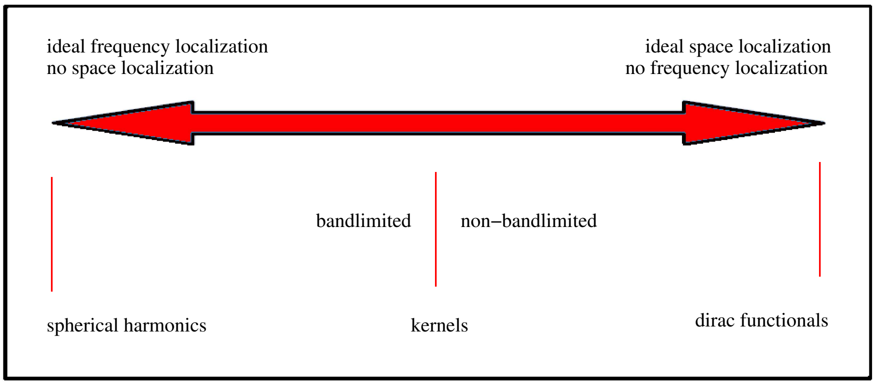


Figure 3.1: Uncertainty principle (Freeden 1999)

for which frequencies above a certain degree are zero. The non-bandlimited versions exhibit a stronger space localization than their respective bandlimited counterparts.

The kernel functions to be discussed in the following are radial basis functions, which implies that they only depend on the spherical distance between a point on the sphere and the respective nodal point at which the kernel is located. Therefore, in the following part of this section, the idea of modeling a function on the sphere in terms of radial basis functions will first be introduced in general. For further information concerning this matter, it can be referred, for example, to Freeden et al. (1998). This general introduction is followed by the specification of the basis functions to be used in the investigations in this thesis. They differ from existing radial basis functions by the particular choice of the coefficients defining the shape of the kernels.

3.4.1 Radial Basis Functions

As mentioned above, radial basis functions $\Phi(\mathbf{x}, \mathbf{x}_i)$ depend only on the spherical distance between two points on the sphere expressed by the unit vectors $\mathbf{x}$ and $\mathbf{x}_i$. Thus they can be regarded as

$$\Phi_i(\mathbf{x}, \mathbf{x}_i) = \Phi(\mathbf{x} \cdot \mathbf{x}_i), \tag{3.72}$$

with $\mathbf{x}, \mathbf{x}_i \in \Omega$. Any radial symmetric basis function can be expressed by a sum of Legendre polynomials,

$$\Phi_i(\mathbf{x}, \mathbf{x}_i) = \sum_{n=0}^{\infty} \sqrt{2n+1} \cdot k_n P_n(\mathbf{x} \cdot \mathbf{x}_i) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} k_n Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{x}_i). \tag{3.73}$$

The coefficients k_n define the shape of the function; the question concerning the choice of these coefficients will be addressed in more detail below. Any functional on the sphere can be modeled as a series expansion of the radial basis functions according to

$$s(\mathbf{x}) = \sum_{i=1}^{I} a_i \Phi_i(\mathbf{x}, \mathbf{x}_i), \tag{3.74}$$

with the scaling coefficients a_i. $\mathbf{x}$ stands for the evaluation point, and the $\mathbf{x}_i$ denote the nodal points at which the basis functions are located on the surface of the sphere. As the gravitational potential is a harmonic function, the radial basis functions are chosen as harmonic kernel functions as well, meaning that they can be outward continued harmonically. In later calculations the gravity field functionals will be modeled on a reference sphere Ω_R with radius R; the upward continuation into the exterior of Ω_R can be performed by applying the upward continuation operator defined in Eq. (3.70),

$$\Phi_i(\mathbf{x}_r, \mathbf{x}_i) = \int_{\Omega} U(\mathbf{x}, \mathbf{x}_r) \Phi_i(\mathbf{x}, \mathbf{x}_i) d\Omega = \sum_{n=0}^{\infty} \frac{1}{R} \left(\frac{R}{r}\right)^{n+1} \sqrt{2n+1}\, k_n P_n \left(\frac{\mathbf{x}_r}{r} \cdot \mathbf{x}_i\right). \tag{3.75}$$

Details concerning harmonic kernel functions can be found in Moritz (1980) or Lelgemann (1981), and further information is given in Freeden et al. (1998) andFreeden (1999).

3.4.2 Spherical Splines

The term 'spline' was originally derived from the denomination of a thin bar used in the geometrical design of ship hulls. This bar assumes a form that minimizes its (one-dimensional) bending energy (Schoenberg 1964a and Schoenberg 1964b). Following this definition, splines are functions that minimize the norm of the mth derivative of a function, $\int_a^b \left(f^{(m)}(\mathbf{x})\right)^2 d\mathbf{x}$; in the case $m = 2$ the minimization of the bending energy mentioned above is achieved, see also Wahba (1990). Transferring the one-dimensional concept of a thin bar to the sphere, it can be visualized as the bending energy of a thin plate covering the surface of the sphere. Concerning the idea of spherical spline approximation, refer to Wahba (1981) and Freeden (1981). The minimization of certain derivatives imposes particular smoothness conditions on the function. This leads to the expansion of the concept of splines to the minimization of norms imposed by reproducing kernels (see, for example, Freeden 1999). Concerning the relationship between smoothness defined by the idea of differentiability and smoothness induced by reproducing kernels, refer to Section 3.3.4. Therefore, the use of the reproducing kernel Φ of a reproducing kernel Hilbert space $\mathcal{H}_\Phi$ as basis functions in Eq. (3.74) can be specified as spherical spline representation of the signal $s(\mathbf{x})$, see also Kusche (2002). It can be distinguished between spline interpolation and spline approximation. In case of interpolation, the values at N observation points are reconstructed exactly by the use of N basis functions, one positioned at each observation point. In contrast to this, spline approximation refers to modeling the (error affected) observation values as accurately as possible (in the least squares sense) by a limited number $I < N$ of basis functions positioned at previously fixed nodal points. These nodal points are independent of the observation points and generally do not coincide with them. The latter case of spline approximation is the one relevant in the context of regional gravity field modeling. I basis functions located at the nodes of a spherical grid are applied to represent the gravity field solution obtained from a large amount of observations given at the satellite's positions. An overview concerning spline representations of functions on the sphere with application in geopotential modeling can, for example, be found in Jekeli (2005).

When dealing with radial basis functions, the coefficients k_n in Eq. (3.73) are responsible for the shape of the basis functions, with different choices for k_n leading to different functions. The coefficients to be selected for the design of the spline functions introduced in this thesis are derived from the covariance function given in Eq. (3.58),

$$\Phi_i(\mathbf{x}, \mathbf{x}_i) = \sum_{n=2}^{\infty} \sigma_n P_n(\mathbf{x} \cdot \mathbf{x}_i) = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{x}_i). \tag{3.76}$$

This means that the coefficients k_n in Eq. (3.73) are chosen according to

$$k_n = \frac{\sigma_n}{\sqrt{2n+1}}. \tag{3.77}$$

A comparison with Eq. (3.58) reveals that k_n are the square roots of the coefficients defining the covariance function. Investigations about this choice of coefficients will be given below in this section, as well as in Section 3.4.5 and in Section 5.3. As was the case for the covariance function, again the coefficients of degree zero and degree one are omitted. The space localizing basis functions are applied to model residual fields, as will be described in detail below. Therefore, it is sufficient for them to model a space of functions for which the zeroth and the first moments vanish, as defined by (3.66).

The basis functions Φ_i can again be used to construct a reproducing kernel Hilbert space $\mathcal{H}_\Phi$ with Φ as a reproducing kernel. The eigenvalues of the kernel are the k_n of Eq. (3.77). This leads, under consideration of Eq. (3.60), to the norm in the space $\mathcal{H}_\Phi$ being defined by

$$\|f(\mathbf{x})\|_\Phi^2 = \sum_{n=2}^{\infty} \frac{\sqrt{2n+1}}{\sigma_n} \sum_{m=-n}^{n} c_{nm}^2. \tag{3.78}$$

If $f(\mathbf{x})$ is the function (e.g. the gravitational potential) from which the σ_n are derived, this leads to the following norm,

$$\|f(\mathbf{x})\|_\Phi^2 = \sum_{n=2}^{\infty} \frac{\sqrt{2n+1}}{\sigma_n} \sigma_n^2 = \sum_{n=2}^{\infty} \sqrt{2n+1}\, \sigma_n. \tag{3.79}$$

The question arises which implications are induced by the choice of the coefficients k_n. They define the shape of the spline kernel and thus the norm of the associated Hilbert space. While a variety of different norms is imaginable, it is reasonable to choose the norm closely adapted to the specific characteristics of the data, e.g. a norm in accordance with a-priori information about the data. Investigating (3.79) reveals that when assuming Kaula's rule (3.45) for the statistical properties of the gravitational potential on the sphere, the terms in the series expansion of the norm would behave according to

$$\sqrt{2n+1} \cdot \sigma_n \sim \frac{1}{n}. \tag{3.80}$$

Thus it can be concluded that the gravitational potential must be assumed to be slightly smoother than Kaula's rule to become part of the Hilbert space $\mathcal{H}_\Phi$. This can be realized from the fact that with $\sigma_n^2 \sim \frac{1}{n^3}$ (as proposed by Kaula) Eq. (3.79) results in a summation of terms of the order $1/n$, which is exactly the limit for which a series does not yet converge. As soon as the function is only slightly smoother, i.e. as soon as the following relationship is given for ϵ being an arbitrarily small positive value,

$$\sum_{n=0}^{\infty} \frac{1}{n^{(1+\epsilon)}} \leq \infty \tag{3.81}$$

convergence is guaranteed. This does not pose a problem, however, as Kaula's rule is only an abstract approximation and does not have to be valid especially for high degrees. An equivalent assumption has to be made for the gravitational potential to have a first derivative in $\mathcal{L}_2$ on the sphere. This is shown for the first radial derivative in an overview dealing with convergence issues given in Section 3.4.5.

3.4.3 Bandlimited Spline Functions

So far, the discussion of the basis functions has been related to infinite dimensional Hilbert spaces. As soon as one deals with practical calculations, however, the series expansion of the basis functions has to be limited to a maximum degree N. This leads to a modification of Eq. (3.73),

$$\Phi_i(\mathbf{x}, \mathbf{x}_i) = \sum_{n=2}^{N} \sqrt{2n+1}\, k_n P_n(\mathbf{x} \cdot \mathbf{x}_i) = \sum_{n=2}^{N} \sum_{m=-n}^{n} k_n Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{x}_i). \tag{3.82}$$

The restriction to a maximum degree corresponds to the omission of higher frequencies. Thus the basis functions can only model a finite dimensional Hilbert space spanned by the Y_{nm} up to degree N, which results in a smoother version of the modeled function. When dealing with bandlimited spline kernels, convergence issues, as given by (3.79) and as will be discussed in Section 3.4.5, are of minor importance, as an infinite norm is impossible in finite dimensional function spaces. The reasonable upper degree for the truncation of the series expansion depends on the expected gravity field signal to be modeled by the corresponding basis functions.

In addition to the omission of the high frequencies, the basis functions to be used in actual regional gravity field recovery problems are always applied to the parameterization of residual fields. Therefore, they are supposed to model information additional to a given global reference model. That is why the basis functions have to be selected, so that they reflect only the spectral characteristics of the residual field. This results in coefficients k_n to be derived from some kind of difference degree variances,

$$k_n = \frac{\Delta\sigma_n}{\sqrt{2n+1}} \qquad \text{with} \qquad \Delta\sigma_n^2 = \sum_{m=-n}^{n} \Delta c_{nm}^2. \tag{3.83}$$

The difference degree variances can either be obtained by actually subtracting the coefficients of the reference field from alternative global models which describe a-priori well-known high-resolution gravity field features. More appropriate, however, is the use of the formal errors of the coefficients of the reference model. They reflect the residual signal information for each frequency still present in the data, but not yet modeled by the reference field. For degrees higher than the maximum degree provided by the reference field, the difference degree variances become the degree variances themselves or an approximation thereof, e.g. provided by Kaula's rule. In Fig. 3.2 examples of (normalized) spline kernels for different resolutions are displayed for a maximum degree of $N = 60$ (top), $N = 120$ (middle), and $N = 240$ (bottom), respectively. They are calculated according to Eq. (3.82) and each of the basis functions is derived from the error degree variances of the GRACE gravity solution ITG-Grace02s (Mayer-Gürr 2006). In case of the spline kernel developed until $N = 240$, the coefficients have been padded with degree variances approximated by Kaula's rule above degree $n = 120$. The illustrations reveal how the kernels become more space localizing with higher expansion degree, because their support becomes increasingly narrow. This shows how the maximum degree accounts for the correlation length of the kernel. However, the bandlimited basis functions are not rigorously space localizing, as it is indicated by the oscillations in Fig. 3.2. Strictly speaking, they have to be addressed as functions with global support, because they are not exactly zero outside a certain support area. But as they decay sufficiently fast, this fact can be neglected, and the spline functions can be referred to as space localizing basis functions despite these minor simplifications. Even if the functions were non-bandlimited, and thus the expansion in Eq. (3.76) was performed up to infinity, the functions would converge for $n \to \infty$. This is because the coefficients k_n can be assumed to decay sufficiently fast to assure that the sum in Eq. (3.76) is finite. With an assumed frequency spectrum given by Kaula's rule (3.45), they behave according to

$$\frac{\Delta\sigma_n}{\sqrt{2n+1}} \sim \frac{1}{n^2}. \tag{3.84}$$

3.4.4 Arrangement of the Basis Functions on the Sphere

Directly connected to the question of the shape of the basis functions, there arises the problem of the arrangement of the spline kernels on the sphere being responsible for the distance between the nodal points of neighboring basis functions. Principally, the radial basis functions are to be distributed as homogeneously as possible on the surface of the sphere. Therefore, a spherical grid is designed, and the spline kernels are located at the nodes of this grid. Besides the uniform distribution, the grid is to meet the requirements of being reproducible and easily implementable on the one hand and of providing the opportunity to chose the number of grid points as flexibly as possible on the other hand. The latter point is important as the number of spline kernels defines the resolution of the gravity field solution. In Section 3.5 several possible distributions are introduced. The so-called triangle vertex grid has turned out to be the most feasible one. It is described in Section 3.5.1.6, and its choice is justified in Section 3.5.2.

The given grid can be defined for different levels of resolution having to be adopted to the width of the basis functions to ensure a reasonable coverage. When, for example, spline kernels with narrow support are located too far apart, no smooth modeling is possible. In the calculations presented in this thesis, the number of global basis functions has been adopted to the maximum expansion degree N of the spline kernel via the resolution of a corresponding spherical harmonic expansion of the same degree N. Concerning the discussion of the comparison between resolution defined for spherical harmonics on the one hand and for space localizing basis functions on the other hand, refer to Section 3.5.3. The number of basis functions, therefore, equals the number of unknown parameters of the corresponding spherical harmonic expansion. This implies that the $(N+1)^2$ spherical harmonic coefficients lead to the same number of projected spline kernels. Due to the limitations in the flexible selection of the number of grid points for the chosen spherical grid, this number can generally not be encountered exactly. This leads to slightly more parameters for the spline representation. Fig. 3.3 shows an example of three adjoining spline kernels developed up to $N = 120$ from GRACE error degree variances and thus being equal to the one displayed in the middle of Fig. 3.2. The resulting grid with a resolution related to $I = 121^2$ basis functions provides a resolution with a spherical distance between the kernels of about 1.67°.

It has to be pointed out that the above considerations are valid when a global homogenous coverage with basis functions is aspired. But the regional recovery approach would also offer the opportunity to adapt the resolution of the gravity field according to the signal content in different regions. This would lead to varying

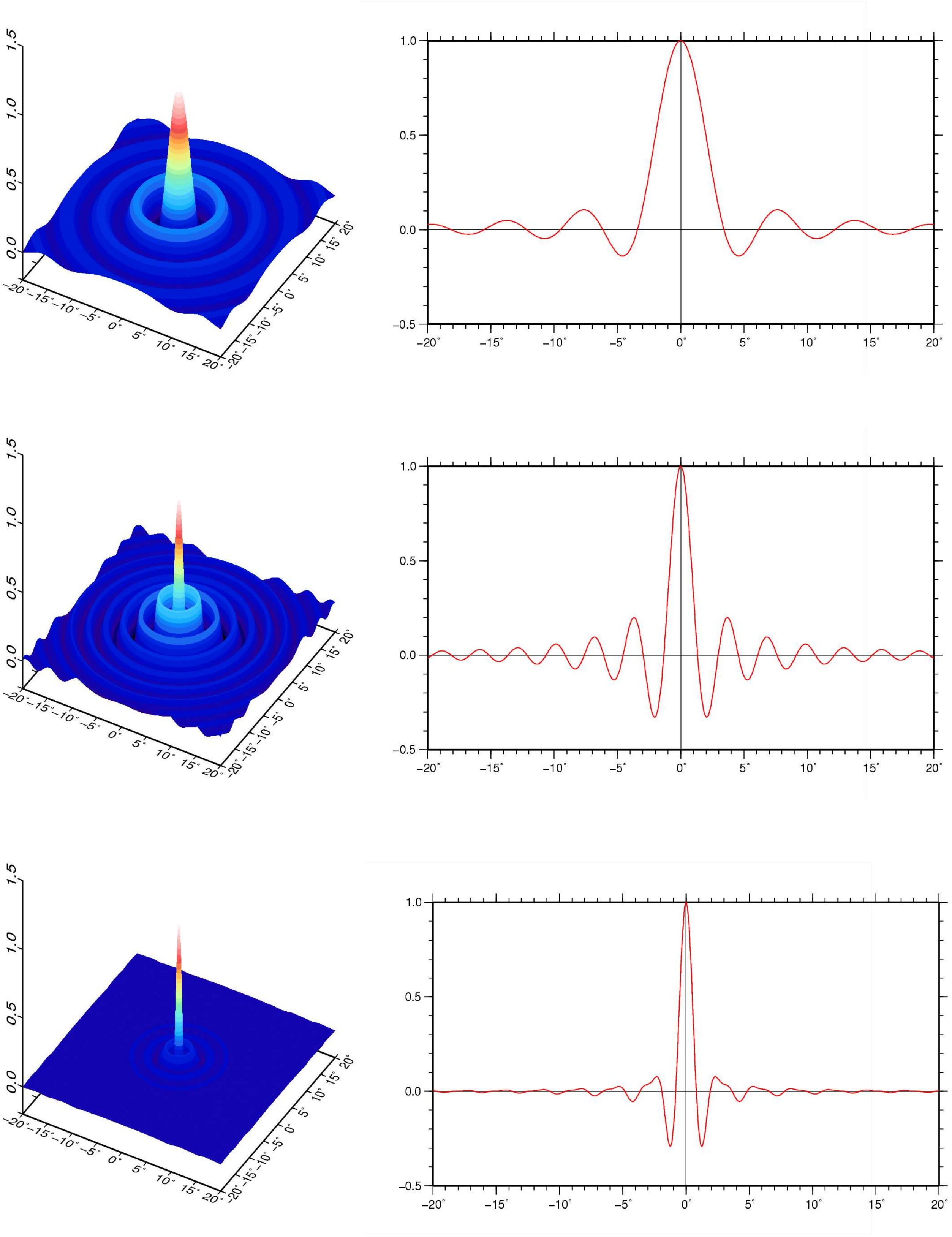


Figure 3.2: Basis functions for different resolutions, top: $N = 60$, middle: $N = 120$, bottom: $N = 240$, shape coefficients derived from error degree variances of ITG-Grace01s, above $n = 120$ padded by Kaula's rule

spline kernels and corresponding grids in different geographical areas. In this case the point distribution has to be defined for each region individually. Furthermore, it is thinkable to use the space localizing basis functions to model particular geophysical phenomena. If so, the basis functions could be arranged in a way to best reflect the characteristics of the particular phenomenon.

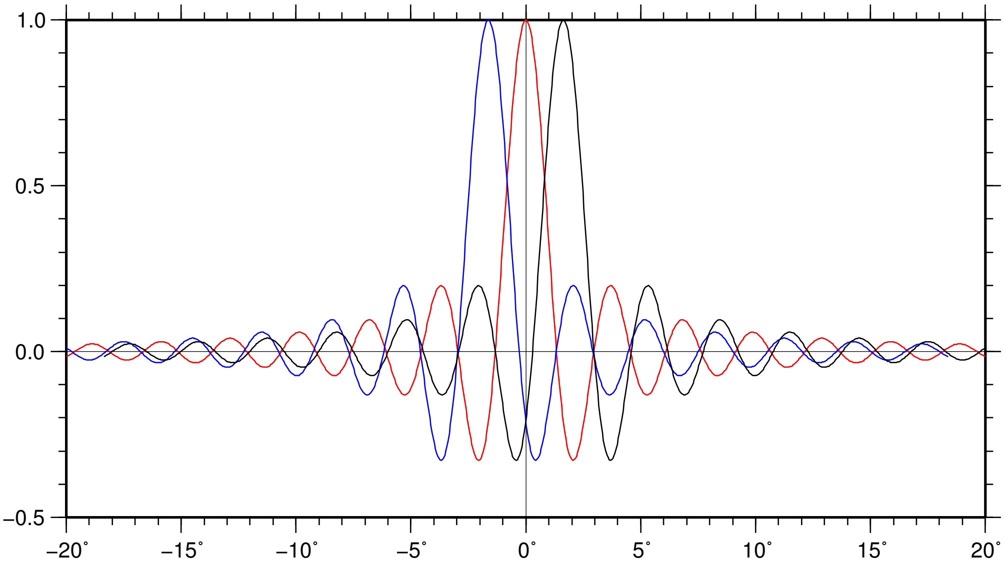


Figure 3.3: Basis functions $N = 120$, 1.67° distance of nodal points

3.4.5 Convergence Issues: Summary

In different sections throughout this thesis, the convergence problem with respect to the norm of different (reproducing kernel) Hilbert spaces is addressed. The question of convergence of the norm is thereby equivalent to the question whether a function belongs to the respective Hilbert space, or whether it does not satisfy the smoothness requirements imposed by the corresponding reproducing kernel. To give an overview of the different Hilbert spaces and to simplify comparisons, the different convergence issues are summarized in the following. Customarily, Kaula's rule of thumb is assumed as approximation of the power spectral density of the gravity field; therefore it serves as basis for some of the following considerations. As stated in Eq. (3.45), Kaula's rule implies that the degree variances of the gravitational potential and the corresponding standard deviations behave according to

$$\sigma_n^2 \sim \frac{1}{n^3} \qquad \text{(Kaula's rule)} \qquad \Rightarrow \sigma_n \sim \frac{1}{n\sqrt{n}}. \tag{3.85}$$

For the gravitational potential modeled as a function f according to (3.38), it is easy to show that its norm converges in $\mathcal{L}_2$,

$$\|f(\mathbf{x})\|^2 = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=0}^{\infty} \sigma_n^2 \qquad < \infty. \tag{3.86}$$

In Eq. (3.64) a condition has been specified to guarantee that not only the function itself belongs to $\mathcal{L}_2$, but that its first radial derivative is part of $\mathcal{L}_2$ as well. Assuming Kaula's rule this implies

$$\left\| \left. \frac{df(\mathbf{x}_r)}{dr} \right|_R \right\|^2 = \sum_{n=0}^{\infty} (n+1)^2 \sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=0}^{\infty} (n+1)^2 \cdot \sigma_n^2 \qquad \text{with} \qquad (n+1)^2 \cdot \sigma_n^2 \sim \frac{1}{n}. \tag{3.87}$$

This relationship between the potential and its first radial derivative is part of the so-called Meissl scheme (Meissl 1971), which also establishes the eigenvalue connections to the second radial derivative and the upward or downward continuation operator. Rummel and van Gelderen (1995) extended the scheme to the first and second order horizontal derivatives and to mixed horizontal-vertical derivatives, deducing similar relationships for the behavior of these eigenvalues. From (3.87) it follows that the first (radial) derivative of a function f belongs to $\mathcal{L}_2$ if $\sigma_n^2 < \frac{1}{n^3}$, i.e. if the gravitational potential is assumed slightly smoother than proposed by (3.85). Concerning the limit of convergence see Eq. (3.81).

To investigate further smoothness properties, it can be examined whether the potential is part of different RKHS. A quite common choice of the reproducing kernel in geodetic applications is the covariance function $\mathcal{C}$ of the potential. It was already shown in Section 3.3.3 that the potential is not part of this RKHS $\mathcal{H}_{\mathcal{C}}$, as it becomes obvious by the relation

$$\|f(\mathbf{x})\|_{\mathcal{C}}^2 = \sum_{n=2}^{\infty} \frac{2n+1}{\sigma_n^2} \sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=2}^{\infty} \frac{2n+1}{\sigma_n^2} \cdot \sigma_n^2 = \sum_{n=2}^{\infty} (2n+1) \qquad = \infty. \tag{3.88}$$

For the design of the spline kernel Φ to be applied as basis functions in the regional gravity field recovery process, a different norm was introduced in Eq. (3.79). Defining an RKHS $\mathcal{H}_{\Phi}$, this norm is given by

$$\|f(\mathbf{x})\|_{\Phi}^2 = \sum_{n=2}^{\infty} \frac{\sqrt{2n+1}}{\sigma_n} \sum_{m=-n}^{n} c_{nm}^2 = \sum_{n=2}^{\infty} \frac{\sqrt{2n+1}}{\sigma_n} \cdot \sigma_n^2 = \sum_{n=2}^{\infty} \sqrt{2n+1} \cdot \sigma_n \qquad \text{with} \qquad \sqrt{2n+1} \cdot \sigma_n \sim \frac{1}{n}. \tag{3.89}$$

This norm converges if $\sigma_n^2 < \frac{1}{n^3}$, as in this case the requirement (3.81) is fulfilled. Comparing Eq. (3.89) and Eq. (3.87), it can be concluded that, for a function f to belong to $\mathcal{H}_{\Phi}$, the same demand has to be made for the smoothness of f as is necessary to guarantee its radial derivative to be part of $\sigma_n^2 < \frac{1}{n^3}$. For both requirements the potential has to be slightly smoother than assumed by Kaula's rule. This does not cause a problem, however, as Kaula's rule only represents an abstract approximation of the frequency behavior exhibited by the degree variances.

Comparing Eq. (3.88) and (3.89), it can be concluded that the choice of the square root of the coefficients of the covariance functions as shape coefficients in the design of the space localizing basis functions results in a reproducing kernel for which the norm of the gravitational potential (almost) converges on the sphere. This would not be the case if the covariance function itself was chosen as basis function.

It shall be mentioned that the investigations of convergence described above were performed on the unit sphere in order to investigate the mathematical properties of the radial basis functions and the implications made by different choices of reproducing kernels. In the outer space, the gravitational potential as harmonic function and all its derivatives have to converge. This is the reason why the Bjerhammer sphere with radius R_B inside of the Earth was introduced, resulting in the introduction of a factor $s = \frac{R_B}{R}$ (with R denoting the mean Earth radius) into the covariance function, see, for example, Tscherning and Rapp (1974).

In the context of regularization (Section 5.2.3), there occurs the norm of the basis function Φ (as a function of $\mathbf{x}$) in an RKHS defined by the covariance function $\mathcal{C}$. This norm is given by

$$\|\Phi(\mathbf{x}, \mathbf{x}_i)\|_{\mathcal{C}}^2 = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \frac{2n+1}{\sigma_n^2} \left(\frac{\sigma_n}{\sqrt{2n+1}} \right)^2 = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} 1 = \infty. \tag{3.90}$$

Obviously, this norm does not converge, a matter that will be discussed in Section 5.2.3.

3.5 Point Distributions on the Sphere

This section deals with different distributions of points on the surface of the sphere. In the context of regional gravity field recovery as treated in this thesis, the problem of choosing the appropriate arrangement of points on the sphere occurs at two different occasions.

- In order to model a regional gravity field refinement, a grid of points has to be created as nodal points for the distribution of space localizing basis functions.
- The combination of different regional gravity field models in order to obtain a global spherical harmonic solution is performed by means of quadrature methods, as described in Section 6.2. In this context the nodes of the numerical quadrature method also have to be located on a specific grid on the sphere.

For both tasks the respective system of points has to meet different requirements. In case of the nodal point pattern for the location of the spline functions, the grid points are to be as homogeneously distributed as possible. Concerning the numerical quadrature, the quadrature nodes and their corresponding quadrature weights (generally associated with the area assigned to the grid points) have to meet the task of calculating spherical harmonic coefficients from the gravity field functionals, evaluated at the grid points, as accurately as possible. In the first part of this section, several grids on the sphere are introduced, and their point distributions are described. Subsequently, those grids have to be examined focussing on how well they satisfy the requirements stated above. Since not all of the described point distributions are suitable for both tasks, the grids are only examined regarding the demands they are intended to fulfill. The investigations concerning the applicability of the uniform grids to suit as nodal point pattern for spline functions are carried out in the second part of this section. For the study of the different grid types as nodes for respective quadrature formulas, refer to Section 6.2.1.

3.5.1 Grids

In the following, different specifications of point distributions are introduced. This is, by no means, a complete discussion of all different possible ensembles of points on the sphere, as such an investigation would be far beyond the scope of this thesis. The discussion is limited to seven grids that are either particularly widely used or especially suitable for the given tasks. Each of the described point arrangements is given as set of coordinates (λ_i, ϑ_j) on the surface of the unit sphere.

3.5.1.1 Geographical Grid

The term 'geographical grid' is used for an equal-angular point distribution with points located along meridians and along circles of latitude. The angular difference between adjacent points along meridians equals the angular difference between adjacent points along circles of latitude. With L denoting the number of parallels, this implies

$$\Delta\lambda = \Delta\vartheta = \frac{\pi}{L}, \tag{3.91}$$

resulting in the following point setting (see Fig. 3.7),

$$\lambda_i = \frac{\Delta\lambda}{2} + i \cdot \Delta\lambda \qquad \text{with} \qquad 0 \leq i < 2L, \tag{3.92}$$

$$\vartheta_j = \frac{\Delta\vartheta}{2} + j \cdot \Delta\vartheta \qquad \text{with} \qquad 0 \leq j < L. \tag{3.93}$$

The number of grid points along the parallels is twice the amount of the number of points along a meridian. This results in a total quantity of points determined by

$$I = 2 \cdot L^2. \tag{3.94}$$

3.5.1.2 Gauss - Grid

The arrangement of points on the sphere at the nodes of a Gaussian grid is strongly connected to the Gauss-Legendre quadrature, as described in Section 6.2.1.2. In this context, it is widely known in numerical analysis (refer to, e.g., LANCZOS 1956) and has been applied to geodesy as well, (see, for example, PAYNE 1971). The grid features equiangular spacing along L circles of latitude with

$$\Delta\lambda = \frac{\pi}{L} \qquad \Rightarrow \qquad \lambda_i = \frac{\Delta\lambda}{2} + i \cdot \Delta\lambda \qquad \text{with} \qquad 0 \leq i < 2L. \tag{3.95}$$

Along the meridians the points are located at L parallels at the L zeros ϑ_j of the Legendre polynomial of degree L,

$$P_L(\cos\vartheta_j) = 0. \tag{3.96}$$

Consequently, the number of grid points sums up to

$$I = 2 \cdot L^2. \tag{3.97}$$

The Gauss grid looks quite similar to the corresponding geographical grid with the same number of parallels. Its distinctive feature is the unique choice of the location of the circles of latitude (Fig. 3.8).

3.5.1.3 Driscoll - Healy Grid

The Driscoll-Healy grid, as introduced by DRISCOLL and HEALY (1994), has equiangular spacing along the meridians as well as along the circles of latitude. In longitudinal direction (along the parallels), these angular differences for a given dimension L coincide with those described for the corresponding geographical grid and Gauss grid. Along the meridians, the size of the latitudinal differences is half the size compared to the geographical grid. This results in the following point pattern (Fig. 3.7),

$$\begin{aligned} \Delta\lambda &= \frac{\pi}{L} & \Rightarrow \qquad & \lambda_i = \frac{\Delta\lambda}{2} + i \cdot \Delta\lambda & \text{with} \qquad & 0 \leq i < 2L, \\ \Delta\vartheta &= \frac{\pi}{2L} & \Rightarrow \qquad & \vartheta_j = j \cdot \Delta\vartheta & \text{with} \qquad & 1 \leq j \leq 2L. \end{aligned} \tag{3.98}$$

Consequently, the number of grid points is

$$I = 4 \cdot L^2. \tag{3.99}$$

The Driscoll-Healy grid is strongly associated with the corresponding quadrature formula, as will be described in Section 6.2.1.1. The concept of the quadrature formula (and the corresponding grid) goes back to NEUMANN (1838); DRISCOLL and HEALY (1994) have developed closed expressions for the respective quadrature weights, therefore the method and the point distribution will in the following be identified as Driscoll-Healy quadrature and Driscoll-Healy grid.

3.5.1.4 Reuter - Grid

The Reuter grid (REUTER 1982) features equi-distant spacing along the meridians determined by the control parameter γ according to

$$\Delta\vartheta = \frac{\pi}{\gamma} \qquad \Rightarrow \vartheta_j = j\Delta\vartheta, \qquad \text{with} \qquad 1 \leq j \leq \gamma - 1. \tag{3.100}$$

Thus $\gamma+1$ denotes the number of points per meridian, as the two poles are included in the point distribution as well. Along the circles of latitude, the number of grid points decreases with increasing latitude in order to achieve an evenly distributed point pattern. This number is chosen, so that the points along each circle of latitude have the same spherical distance as two adjacent latitudes. The resulting relationship is given by

$$\Delta\vartheta = \arccos\left(\cos^2\vartheta_j + \sin^2\vartheta_j \cos\Delta\lambda_j\right). \tag{3.101}$$

The left hand side of this equation is the spherical distance between adjacent latitudes, the right hand side stands for the spherical distance between two points with the same polar distance ϑ_j and a longitudinal difference of $\Delta\lambda_i$. This longitudinal distance can be adjusted depending on ϑ_j to fulfill Eq. (3.101). The resulting formula for $\Delta\lambda_i$ is

$$\Delta\lambda_j = \arccos\left(\frac{\sin\Delta\vartheta - \cos^2\vartheta_j}{\sin^2\vartheta_j}\right). \tag{3.102}$$

The number of points γ_j for each circle of latitude can then be determined by

$$\gamma_j = \left[\frac{2\pi}{\Delta\lambda_j}\right]. \tag{3.103}$$

Here the Gauss bracket $[x]$ specifies the largest integer equal to or less than x. The longitudes are subsequently determined by

$$\lambda_{ij} = \frac{\Delta\lambda_j}{2} + i\cdot(2\pi/\gamma_j), \qquad \text{with} \qquad 0 \leq i < \gamma_j. \tag{3.104}$$

The number of grid points can be estimated by

$$I =\leq 2 + \frac{4}{\pi}\gamma^2, \tag{3.105}$$

as described, for example, in Freeden et al. (1998). The "$\leq$" results from the fact that the γ_j are restricted to integer values.

Approximately the same grid is generated when using the equal area method as in a similar way described by Rapp (1971) and adapted in Hajela (1973). It constitutes that the area of a surface element of the size $\Delta\lambda_j \times \Delta\vartheta$ located at an arbitrary polar distance ϑ_j is set equal to the area of the respective surface element at the equator,

$$\int\limits_{-\frac{\Delta\lambda}{2}}^{\frac{\Delta\lambda}{2}} \int\limits_{\frac{\pi}{2}-\frac{\Delta\vartheta}{2}}^{\frac{\pi}{2}+\frac{\Delta\vartheta}{2}} \sin\vartheta\, d\vartheta\, d\lambda = \int\limits_{-\frac{\Delta\lambda_j}{2}}^{\frac{\Delta\lambda_j}{2}} \int\limits_{\vartheta_j-\frac{\Delta\vartheta}{2}}^{\vartheta_j+\frac{\Delta\vartheta}{2}} \sin\vartheta\, d\vartheta\, d\lambda. \tag{3.106}$$

Solving both integrals leads to the following relationship,

$$\sin\vartheta_j \Delta\lambda_j = \Delta\lambda. \tag{3.107}$$

Here $\Delta\lambda$ denotes the angular difference along the equator which coincides with the angular difference along the meridians $\Delta\lambda = \Delta\vartheta$. This results in the following angular differences along the circles of latitude as an alternative to (3.102),

$$\Delta\lambda_j = \frac{\Delta\vartheta}{\sin\vartheta_j}. \tag{3.108}$$

Regarding Eq. (3.108), it can easily be observed how the distance between longitudes increases when ϑ approaches either zero or π, and how it reaches its minimum at the equator. Even though the $\Delta\lambda_j$ differ slightly when the results of Eq. (3.102) and Eq. (3.108) are compared, the number of points per circle of latitude γ_j is equal in just about every case after conversion to integer values. Exceptions are the northern and southern most circles of latitude, where the two γ_j may differ by one, as was tested for a variety of different resolutions.

3.5.1.5 Triangle Center Grid

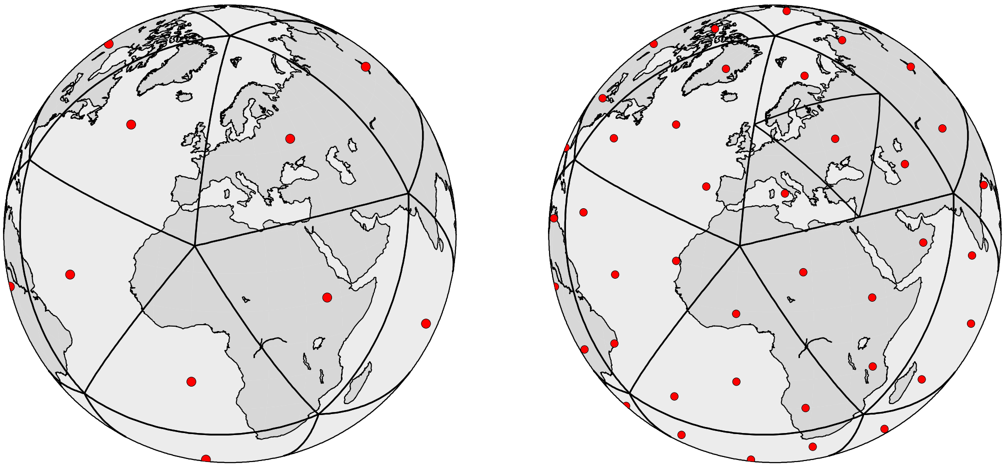
Two different triangular grids will be discussed and evaluated in this chapter. Both, the triangle center mode, as described here, and the triangle vertex mode, as described in the next section, are based on the geometric shape of an icosahedron. The icosahedron is a convex polyhedron with 20 faces, all of them being equilateral triangles, with five of them meeting at any one of the 12 vertices. It represents one of the five platonic solids, which are identified by the fact that they are convex regular polyhedrons, implying that their sides, edges, and angles are all congruent. The icosahedron is the platonic solid with the most possible faces. As the vertices of every platonic solid lie on a sphere, they can be applied as a basis in the construction of point distributions on a sphere. To create triangular point distributions, platonic solids featuring triangles as faces (besides the icosahedron these are the tetrahedron and the octahedron) are best suitable. A triangular grid based on the icosahedron can be created by placing the icosahedron into the sphere with the 12 vertices at the surface of the sphere. For the grids discussed here, one vertex is placed coinciding with the north pole and one with the south pole, although different constellations are possible as well. The surface of the sphere is then divided into 20 equal area spherical triangles by connecting the adjacent vertices of the icosahedron by arcs of great circles. This results in 20 equal area spherical triangles, as displayed on the left side of Fig 3.4. As the icosahedron is the platonic solid with the most (triangular) faces, this arrangement of triangles is the largest number of exactly equal triangles possible to be arranged on a sphere.

In case of the first of the triangle grids, the points of the zeroth level are located at the centers of the icosahedron triangles. Due to this fact, the resulting grid is here denominated by the term 'triangle vertex'. To achieve a finer grid, each of the triangles is divided into four smaller triangles by connecting the midpoints of the triangle edges, as illustrated for one triangle in the right part of Fig 3.4. The refined grid points are again located at the center of the triangles. Subsequently, the triangles can be further densified up to the desired level of densification n. This kind of partitioning is widely used in geodetic or geophysical applications, see, for example, Saff and Kuijlaars (1997), Freeden (1999), Stuhne and Peltier (1999), or Kusche (2002).

The number of grid points for a certain level of densification can be determined by

$$I = 20 \cdot 4^n. \tag{3.109}$$

Thus the quantity of grid points depends exponentially on the level n, as with every additional level the number of grid points quadruplicates.



Icosahedron (level 0)

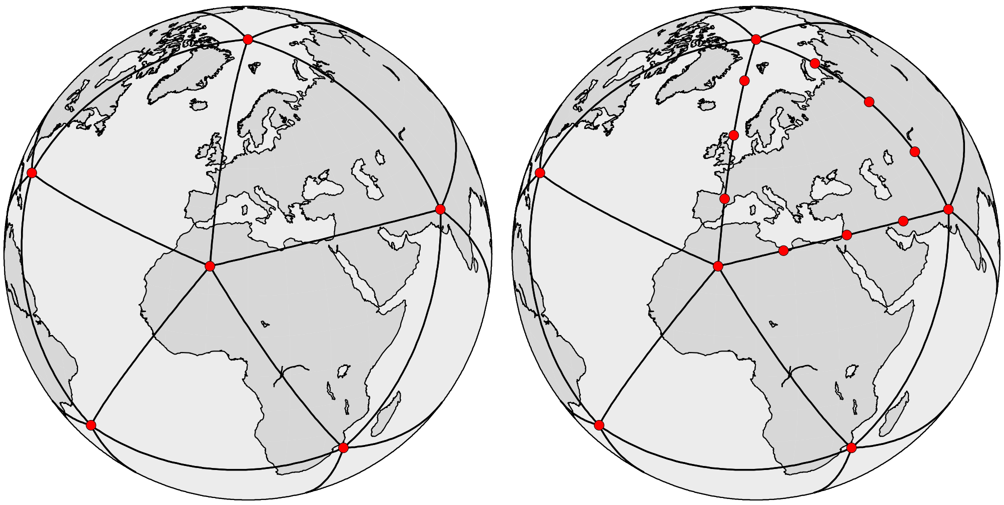
Triangle Center (level 1)

Figure 3.4: Construction of the grid triangle center

3.5.1.6 Triangle Vertex Grid

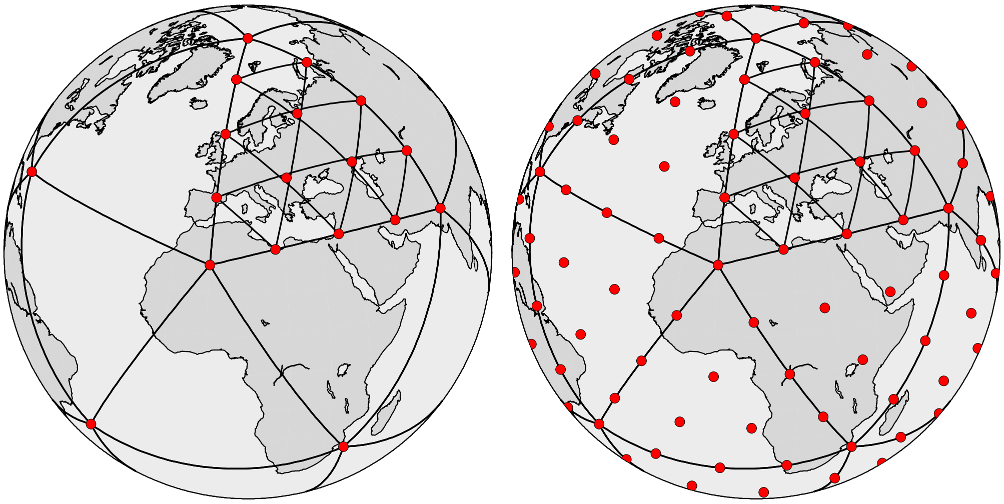
The second point distribution that is based on the icosahedron is referred to as triangle vertex grid, which accounts for the fact that the grid points are not located at the centers of the triangles, but at the vertices. Thus the zeroth level of densification coincides with the 12 icosahedron vertices, as displayed in the upper left part of Fig. 3.5. Then, depending on the envisaged densification, each triangle edge is divided into n parts, illustrated in the upper right part of Fig. 3.5. The new nodes on the edges are then connected by arcs of great circles parallel to the triangle edges. The intersections of each three corresponding parallel lines become nodes of the densified grid as well. As in case of a spherical triangle those three connecting lines do not exactly intersect in one point, the center of the resulting triangle is used as location for the new node (lower left part of Fig. 3.5). The lower right side of Fig. 3.5 finally shows the densified triangle vertex grid for a level of $n = 3$. The number of grid points in dependence of the chosen level of densification can be calculated by

$$I = 10 \cdot (n+1)^2 + 2. \tag{3.110}$$



Icosahedron

Subdivision of the edges



Connecting the points on the edges

Densified grid

Figure 3.5: Construction of the grid triangle vertex, level 3

In the literature dealing with geodetic domes, this kind of partitioning is known as alternate breakdown, see, for example, Kenner (1976). As an example for geodetic applications, Schmidt (1981) was concerned with equidistant sampling point distributions on the sphere for the location of sampling functions. In this context, similar triangular partitionings were mentioned as well.

3.5.1.7 Recursive Quasi Random Grid

This kind of grid distributes an arbitrarily chosen number of I points on the surface of the sphere, following a recursive, quasi random sequence. In longitudinal direction the pattern follows

$$\Delta\lambda = \frac{2\pi}{I} \qquad \Rightarrow \qquad \frac{\Delta\lambda}{2} + \lambda_i = i \cdot \Delta\lambda \qquad \text{with} \qquad 1 \le i \le I. \tag{3.111}$$

This implies that every grid point features a unique longitude, with equi-angular longitudinal differences.

The polar distance in the form $t_i = \cos\vartheta_i$ for each point is determined by the following recursive sequence:

- Starting from an interval $t \in [-1, 1]$.
- If $I = 1$, then the midpoint of the interval is returned as result of the sequence, and the sequence is terminated.
- If the number of points is uneven, the midpoint is included into the list of t_i.
- Subsequently, the interval is bisected into an upper and lower half, and the sequence is called for both halves.
- t from upper and lower half are alternately sorted into the list of t_i.
- The polar distances are calculated by

$$\vartheta_i = \arccos t_i. \tag{3.112}$$

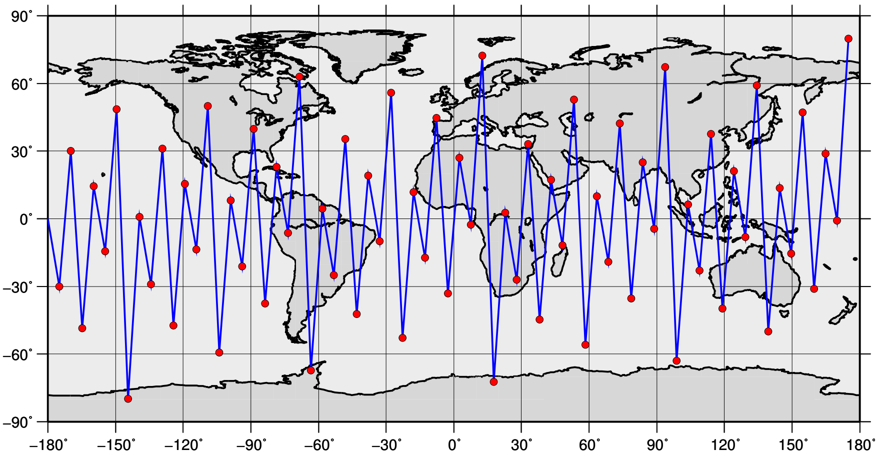


Figure 3.6: Points generated by quasi random grid

The result of the recursive sequence is displayed for an exemplary number of $I = 71$ in Fig. 3.6. It can be clearly observed that points from the lower half of the sequence (southern hemisphere) and from the upper half (northern hemisphere) are taken alternately. And when surveying the bisected sequences (e.g. only the northern hemisphere or bisections thereof), the alternate pattern can be recognized as well.

Alternative sequences can be found, for example, based on van der Corput (1935), e.g. the so-called Hammersley sequence (cf. Niederreiter 1992).

3.5.1.8 Comparison of the Point Distributions

The grids described above are displayed in Fig. 3.7. The resolutions of the different point distributions are chosen such that they provide a comparable point density around the equator. The following specifications of the different grid types have been used:

- Upper left: Geographical grid with $\Delta\lambda = \Delta\vartheta = 3°$
- Upper right: Driscoll-Healey grid of dimension $L = 30$
- Middle left: Triangle vertex grid of level $n = 22$
- Middle Right: Triangle center grid of level $n = 4$
- Lower left: Reuter grid with $\gamma = 64$
- Lower right: Quasi-random grid with $I = 5200$ global points

In the given resolution the Gauss grid is almost indistinguishable from the geographical grid, therefore it is not displayed separately. In order to illustrate the difference between the geographical grid and the Gauss grid, the two grids are compared for a lower resolution in Fig 3.8 to enable the differences to be observed more clearly.

3.5.2 Applicability as Nodal Points for Splines

The space localizing basis functions $\Phi(\mathbf{x}, \mathbf{x}_i)$ are to be distributed evenly over the sphere at the nodes $\mathbf{x}_i$ of some possibly equi-distant grid. Additionally, in order to accommodate the grid according to the desired resolution and the necessary number of unknown parameters, the number of grid points should be adjustable as flexibly as possible. Thus the two requirements for a grid which is applied to build the system of nodal points for the harmonic spline functions are:

- Uniform distribution
- Flexible adjustment of the number of grid points

This section deals with the investigation in how far the described point distributions are suitable to account for the two requirements. Since grids such as the geographical grid, the Gauss grid, or the Driscoll-Healy grid are not capable of providing a uniform distribution at all, they are not taken into account here. Thus the discussion will be limited to the two different triangular grids, the Reuter grid and the pseudo-random grid.

3.5.2.1 Uniform Distribution

The term 'uniformly distributed' or 'evenly distributed' points on a sphere does not have a unique definition. On the contrary, it can be used in various ways, resulting in different criteria which can be introduced to evaluate how even a given point distribution actually is. In a certain sense, only the five Platonic solids achieve regular tesselations, as they are the only ones whose faces are regular and equal, with each vertex incident to the same number of faces. This fact has been widely observed, see, for example, White et al. (1992). But generally, even distribution does not focus so much on the induced tessellation, as it does on the distances and the arrangement of the points/vertices. This leads to a wide variety of criteria to measure the homogeneity of a point distribution. Among these criteria are the following that will be discussed in this section:

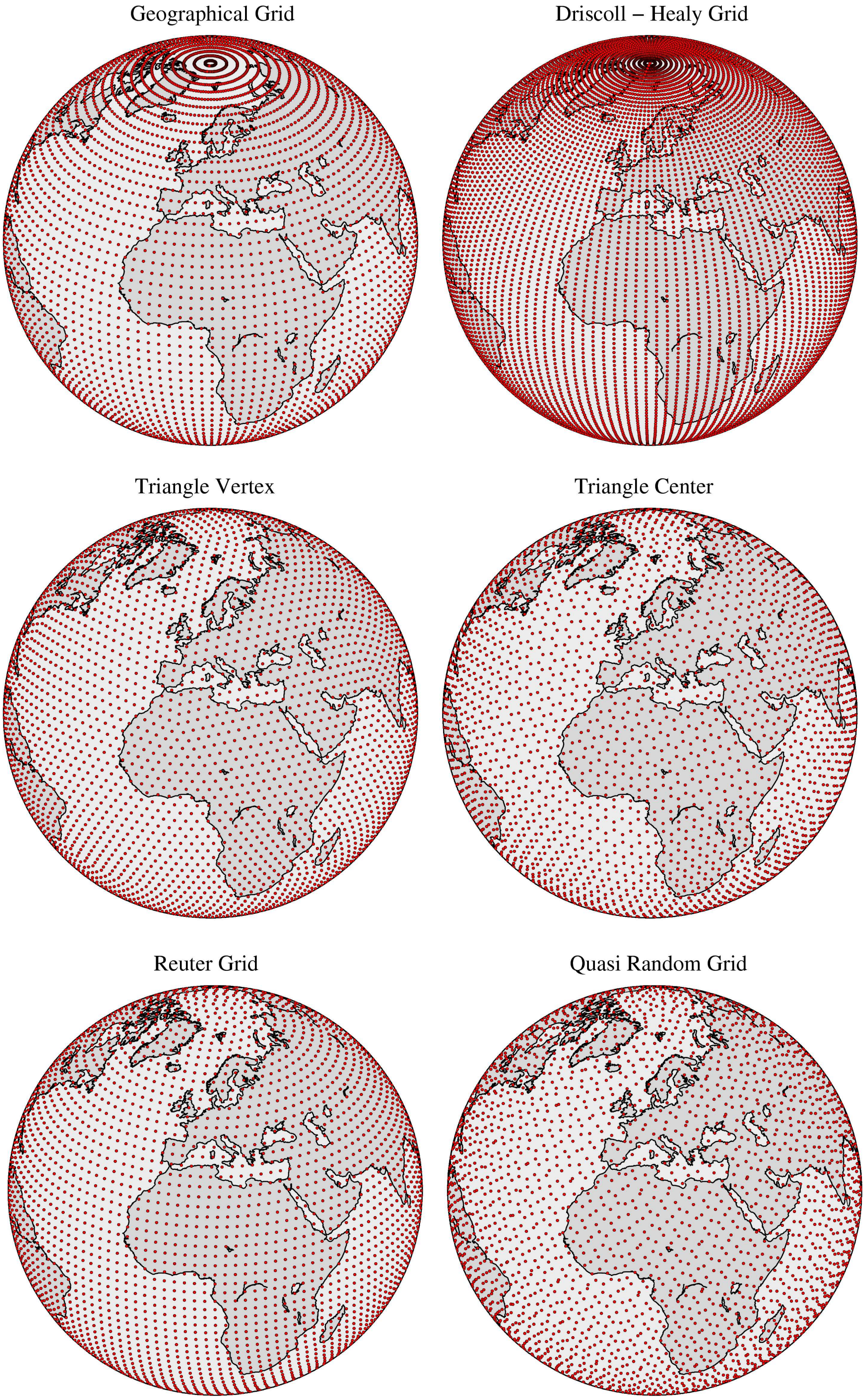


Figure 3.7: Overview of different point distributions on the sphere

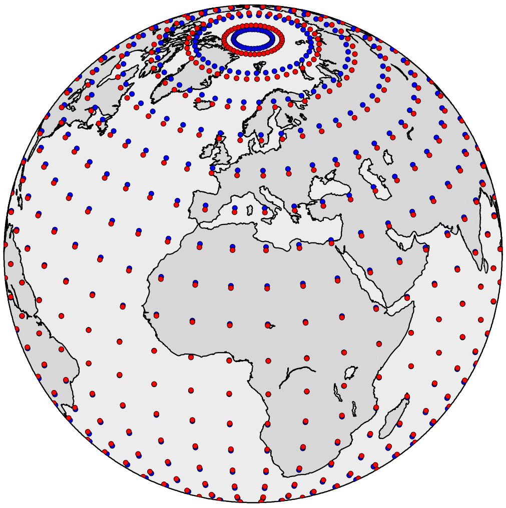


Figure 3.8: Comparison of Gauss grid (red) and geographical grid (blue), L=18

- Equal area partitioning
- Minimizing the potential energy of charged particles at the grid points
- Maximizing the minimal distance between grid points (packing problem)
- Minimizing the maximal distance of any point on the sphere from the closest grid point (covering problem)

There is a variety of additional criteria to measure the evenness of an ensemble of points on a sphere, maximizing the volume of the convex hull can be given as an example. For more information on this topic refer to, e.g., Conway and Sloane (1998).

All these measures can be used as optimization criteria, each leading to slightly different point distributions. The corresponding optimization algorithms modify a given point setting until the chosen criterion is fulfilled as accurately as possible. The difficulty of such optimization problems on the sphere is that they feature a lot of local minima (Conway and Sloane 1998), making the calculations quite complex. Except in a few special cases of numbers of points, it is very hard to prove that a certain arrangement of points is the global minimum. Therefore, a lot of these problems are still unsolved today. Within the scope of this thesis, it is not intended to determine an optimal arrangement for a definite number of points best satisfying any of the criteria. The required optimization algorithms would suffer from the problems described above, and the resulting point distributions would each be valid only for a very specific number of points. For a different resolution, the optimization algorithm would have to be run all over again. Furthermore, the outcome would depend on the abortion criterion defining the convergence of the optimization algorithm. In contrast to this, the possible nodal point pattern should be reproducible and easy to implement, as this is a necessary requirement to enable the use of the spline models for subsequent users. Thus the point setting has to follow a predefined specification. Therefore, the criteria of evenness described above do not serve as optimization criteria, rather will it be investigated how well the point distributions described in Section 3.5.1 fulfill theses requirements. The result is the statement which of them is best suitable to serve as nodal point distribution for the spline functions.

The following investigations will be performed for the two triangle grids, the Reuter grid, and the quasi random grid, as those are the more or less homogeneous ones, as demonstrated in Fig. 3.7. The resolutions of those grids are chosen such that they offer approximately equivalent numbers of grid points. Due to the restrictions concerning the choice of the number of points inflicted by the respective grid specifications, it is not possible to create the different point distributions with an exactly coinciding quantity of points. The following specifications are applied in the investigations:

- Triangle vertex grid, level = 22 => 5292 points
- Triangle center grid, level = 4 => 5120 points
- Reuter, $\gamma = 64$ => 5180 points
- Quasi random grid with 5200 points

Equal Area Partitioning The criterion of equal area partitioning identifies the points on a grid to be the more evenly distributed, the more the sizes of their surface elements agree with each other. In case of grid points being located at the center of distinctively bounded surface elements as, for example, in the geographical grid the determination of the area associated with each grid point is fairly straightforward. This cannot be assumed, however, for all point arrangements dedicated to provide an evenly distributed set of points. Therefore, the surface element Ω_i related to a specific grid point $\mathbf{x}_i$ can more generally be defined as the region of the surface of the sphere being closer to $\mathbf{x}_i$ than to any one of the other grid points,

$$\Omega_i = \{\mathbf{x} | \psi(\mathbf{x}, \mathbf{x}_i) < \psi(\mathbf{x}, \mathbf{x}_j), i \neq j\}. \tag{3.113}$$

To determine this region, the concepts of Delaunay triangulation and Voronoi diagrams have to be introduced.

The Delaunay triangulation (DELAUNAY 1934) of a set of points P is a triangulation $D(P)$ such that no point of P is located within the circumcircle of any triangle of $D(P)$. This requirement is known as circumcircle condition, and its fulfillment simultaneously guarantees the maximization of the minimum interior angle over all triangles. The Delaunay triangulation yields a unique solution as long as there are not more then three points on any one of the circumcircles. Otherwise, there is more than one possible Delaunay triangulation. In this thesis, the Delaunay triangulation is applied in the construction of a Voronoi diagram for arbitrary point distributions on the sphere. Therefore, the triangulation described above has to be applied to the surface of the sphere, calculating spherical triangles between grid points.

The Voronoi diagram is named after the Russian mathematician Georgy Voronoi (VORONOI 1908), even though the same concept was introduced by others as well. The Voronoi diagram $V(P)$ of a set of points P is the partition of the space containing P into cells, with each cell consisting of the points closer to one particular point of P than to any other point of P. The Voronoi diagram of P corresponds to the dual graph of the Delaunay triangulation, meaning that when a Voronoi diagram is determined for the nodes of $V(P)$, then the Delaunay triangulation of the original set of points P is obtained. The Voronoi diagram $V(P)$ can be constructed from the corresponding Delaunay triangulation $D(P)$, as the circumcenters of the triangles of $D(P)$ are the nodes of the Voronoi diagram. The circumcenter can be constructed as the intersection of the perpendicular bisectors of each of the triangle edges. Fig. 3.9 displays the relationship between the Delaunay triangulation and the Voronoi diagram. The black lines denote the triangulation and the blue lines the corresponding Voronoi polygons.

In this thesis, the Voronoi diagram is constructed in order to calculate the area associated with individual grid points for an arbitrary distribution of points on the sphere. This area is equal to the area of the respective Voronoi polygon. For this purpose, in a first step, the Delaunay triangulation is determined. The circumcenter $\mathbf{x}_c$ of a spherical triangle $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ can be calculated by the cross product of the difference vectors between each two of the grid points belonging to the triangle. The normalization takes care of the fact that the circumcenter has to be located on the surface of the unit sphere,

$$\mathbf{x}_c = \frac{(\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{x}_3 - \mathbf{x}_1)}{|(\mathbf{x}_2 - \mathbf{x}_1) \times (\mathbf{x}_3 - \mathbf{x}_1)|}. \tag{3.114}$$

An illustration of a global Delaunay tessellation and the corresponding Voronoi diagram is shown for the grid triangle vertex in Fig. 3.10. To judge how much the areas associated to each grid point vary in different grids, the use of histograms is a useful tool. They graphically evaluate the distribution of the area sizes. The histograms of the four evaluated point distributions are displayed in Fig. 3.11. In addition, the standard deviations of the area sizes are specified in Tab. 3.1. Obviously, in case of the Reuter grid, the grid points

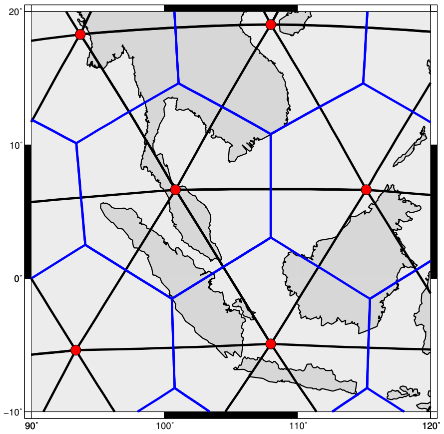


Figure 3.9: Voronoi diagram (blue) and Delaunay triangulation (black)

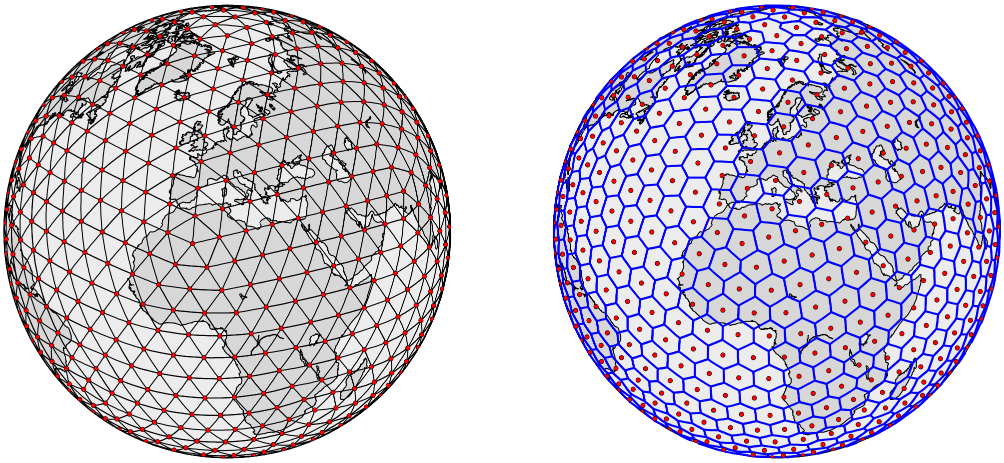


Figure 3.10: Delaunay triangulation (left) and Voronoi diagram (right) of a triangle vertex grid, level 8

grid	σ
Vertex22	0.0041157
Center4	0.0151813
Reuter64	0.0017537
Random5200	0.07275979

Table 3.1: Standard deviation of the area sizes

all have very similarly sized associated areas. This can be observed in the histogram in the upper right part of Fig. 3.11 as well as from the standard deviation given in Tab. 3.1. It is significantly lower compared to the standard deviation of the other point distributions. This appears to be a very straightforward conclusion when considering that the alternative derivation of the Reuter grid, as described by (3.106), is exactly based on the assumption of equal area per grid point. The Reuter grid is followed by the grid triangle vertex whose area sizes are fairly close together as well. This is displayed in the lower left part of Fig. 3.11. Its standard deviation is still considerably smaller than the two remaining options. The grid triangle center and the quasi random grid perform comparably poorly. Regarding the quasi random grid, it has to be pointed out that the histogram even extends over a different scale than the one used for the other distributions.

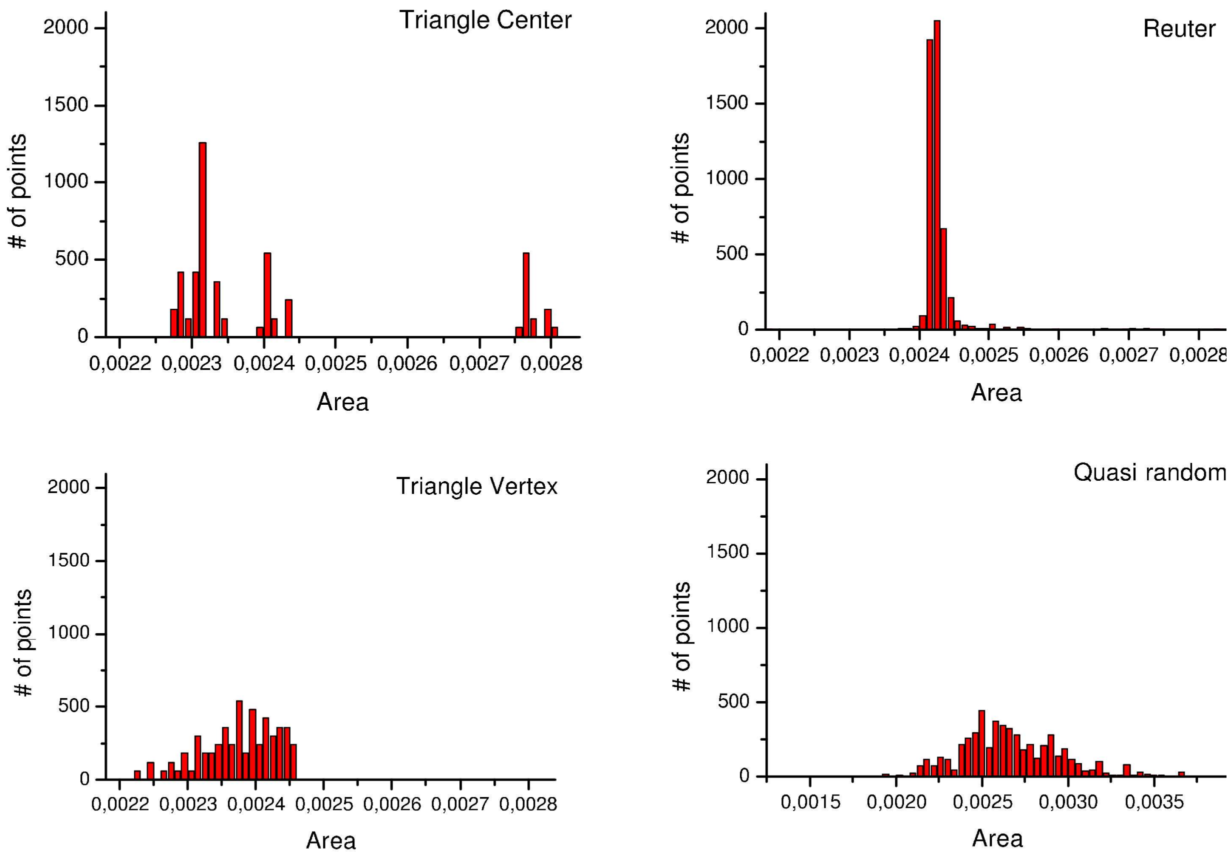


Figure 3.11: Histogram of area sizes associated with individual grid points (different scale for quasi random grid)

Minimal Distance Between Grid Points The problem of determining the minimal distance between two grid points is also referred to as 'packing problem', as it corresponds to the determination of the maximum diameter of I equal spherical caps to be placed on the sphere without overlap. The larger this diameter is, the more evenly the points are distributed. Maximizing the minimal distance between grid points is also known as 'Tammes problem' (Tammes 1930), for further description see, for example, Kottwitz (1991). For each point $\mathbf{x}_i \in P$, the nearest neighbor is identified, and the spherical distance between the neighboring points is calculated,

$$d_{\min_i} = \min_{\mathbf{x}_j \in P, i \neq j} \psi(\mathbf{x}_i, \mathbf{x}_j). \tag{3.115}$$

$\psi(\mathbf{x}_i, \mathbf{x}_j)$ denotes the spherical distance between two points on the surface of the sphere calculated by $\psi(\mathbf{x}_i, \mathbf{x}_j) = \arccos(\mathbf{x}_i \cdot \mathbf{x}_j)$. The minimum of these distances over all points is determined as

$$d_{\min} = \min_{\mathbf{x}_i, \mathbf{x}_j \in P, i \neq j} \psi(\mathbf{x}_i, \mathbf{x}_j), \tag{3.116}$$

and the results are displayed in Tab. 3.2. Here $d_{\min}$ corresponds to the maximum diameter mentioned above, $d_{\min}/2$ is, therefore, also known as packing radius. Here the problem arises that in order to perform a fair comparison, each of the grids would need exactly the same number of points. As this cannot be the case, some kind of normalization has to be applied to account for the differences. Therefore, each distance has to be divided by the square root of the average surface area per point,

$$\bar{d}_{\min} = \frac{d_{\min}}{\sqrt{\frac{4\pi}{I}}}, \tag{3.117}$$

as this corresponds to a division by an optimal distance between neighboring points which should have the order $1/\sqrt{I}$ (Saff and Kuijlaars 1997). Here only one configuration of points for each grid type is investigated, the one with the specific number of points as indicated above. But when applying the normalization (3.117), the value $\bar{d}_{\text{min}}$ stays approximately constant for a respective grid type for a wide variety of point numbers. Therefore, the results obtained for the specific point arrangement can be judged as representative for the respective grid. The Reuter grid and the triangle vertex grid exhibit a very similar minimal point distance which is significantly larger than the one featured by the triangle center grid and the quasi random grid. Especially the quasi random grid does not perform well concerning this criterion, as the recursive sequence may place single points quite close to each other. A closer look is taken on the Reuter and the triangle vertex grid by examining the (normalized) minimum distances $\bar{d}_{\text{min}_i}$ for each point in terms of histograms displayed in Fig. 3.12.

grid	I	$d_{\text{min}}[°]$	$\bar{d}_{\text{min}}$
Vertex22	5292	2.758041	56.5987
Center4	5120	2.143979	43.2762
Reuter64	5180	2.812500	57.1021
Random5200	5200	0.219152	4.4581

Table 3.2: Minimal distances between neighboring grid points

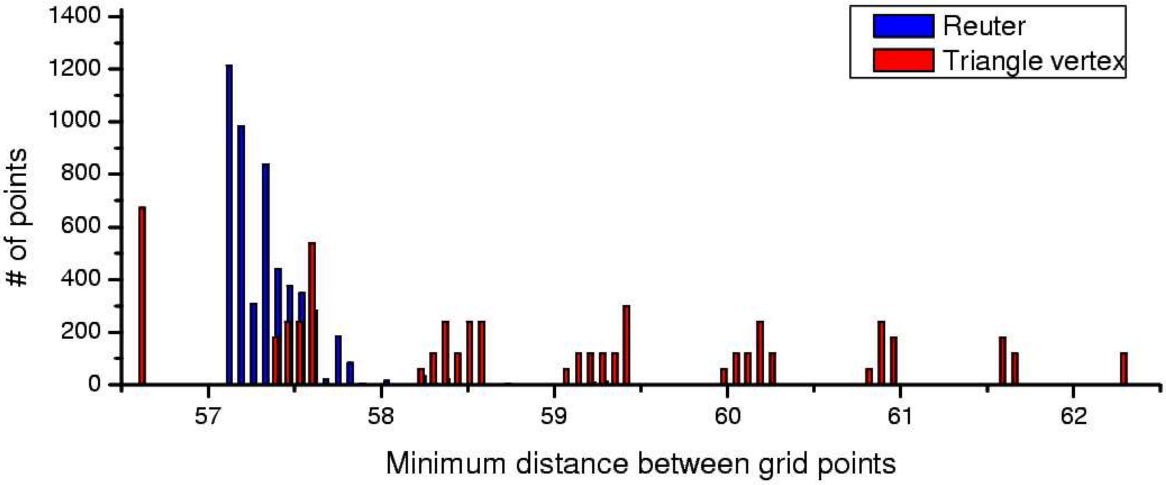


Figure 3.12: Minimal distances between each grid point and the closest of the remaining grid points

Covering Problem Determining the maximum distance from any point on the surface of the sphere to the nearest of the grid points is indicated as 'covering problem' (Conway and Sloane 1998). It describes the smallest radius that equal sized caps (located at the grid points) would need to cover the complete surface of the sphere. In other words, for each surface element Ω_i associated with the grid point $\mathbf{x}_i$ the point $\mathbf{x} \in \Omega_i$ farthest away from $\mathbf{x}_i$ is determined. It satisfies

$$d_{\text{max}_i} = \sup_{\mathbf{x}\in\Omega_i} \psi(\mathbf{x}, \mathbf{x}_i). \tag{3.118}$$

This farthest point has to be one of the nodes of the Voronoi polygon belonging to the grid point $\mathbf{x}_i$. The Voronoi polygon defines the outer bound of the surface element Ω_i, and every point on its edges is closer to the respective grid points than the corresponding Voronoi nodes. The maximum $d_{\max}$ of these maximum distances $d_{\max_i}$ over all points of the grid is given by

$$d_{\max} = \max_{i=1,\ldots,I} \sup_{\mathbf{x}\in\Omega_i} \psi(\mathbf{x}, \mathbf{x}_i). \tag{3.119}$$

This is also referred to as covering radius or mesh norm, and the given set of points is the more evenly distributed, the smaller the maximum distance is. This distance, according to Eq. (3.119), is calculated for the four described grid types, and the results are listed in Tab. 3.3. Again the differences regarding the number of grid points for each point distribution have to be taken into consideration, as mentioned above,

$$\bar{d}_{\max} = \frac{d_{\max}}{\sqrt{\frac{4\pi}{I}}}. \tag{3.120}$$

From Tab. 3.3 it can be concluded that the triangle vertex grid performs significantly better regarding this criterion than the other options, with the Reuter grid being second best. Therefore, a closer look is again taken at these two point distributions, examining the distances $d_{\max_i}$ between each of the grid points and the farthest point of its surface element. The results are displayed in terms of histograms in Fig. 3.13.

grid	I	$d_{\max}[°]$	$\bar{d}_{\max}$
Vertex22	5292	1.7681	36.2828
Center4	5120	2.7153	54.8082
Reuter64	5180	2.2193	45.0586
Random5200	5200	2.5321	51.5086

Table 3.3: Maximum distance of any point on the sphere to its nearest grid point

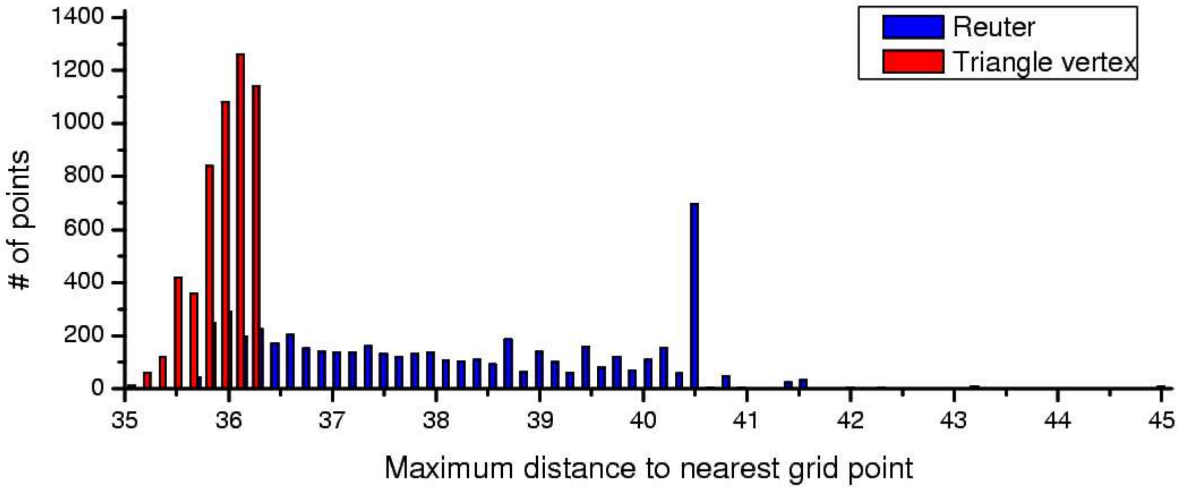


Figure 3.13: Maximum distance of each grid point and the farthest point of its surface element

The impact of the covering problem can be interpreted as follows: For a given number of points a certain average area can be allocated to each grid point $\mathbf{x}_i$. The points in this area would have the shortest distance from $\mathbf{x}_i$ if the surface element Ω_i was a spherical cap with $\mathbf{x}_i$ in its center. Of course, this is not possible in a real point distribution, as the surface of a sphere cannot be segmented into spherical caps. But the maximal distance from any point on the sphere to the nearest grid point is a measure for how far the arrangement of points differs from this ideal situation. Segmenting the surface into spherical caps would also be perfect

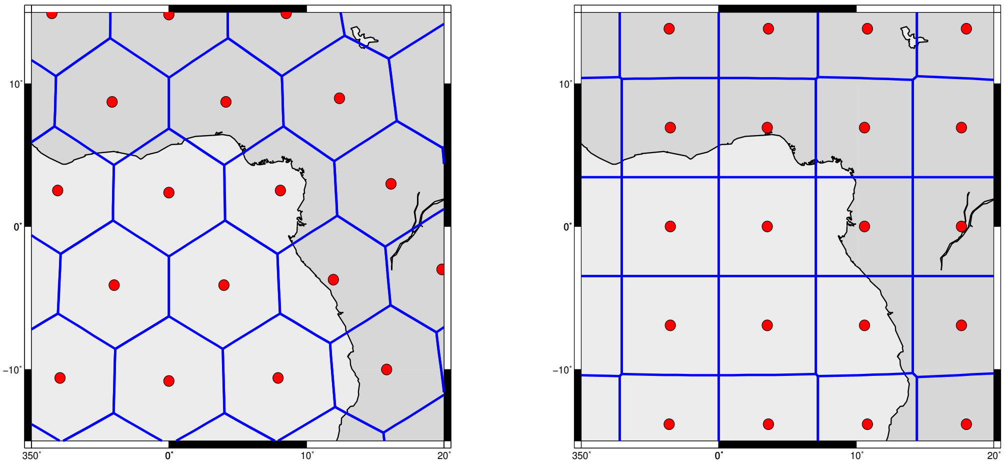


Figure 3.14: Voronoi diagram of triangle vertex grid (left, level: 8) and Reuter grid (right, $\gamma = 25$)

for the arrangements of space localizing basis functions, as they are isotropic and thus represent a somehow 'circular' shape.

The differences between the grid triangle vertex, showing the least maximal distance between arbitrary points and the nearest grid points, and the Reuter grid, performing second best regarding this criterion, can be observed in Fig. 3.14. Here the Voronoi diagram, and thus the surface elements associated with the individual grid points, are displayed. The Voronoi cells in case of the triangle vertex grid with their hexagonal (and for some points pentagonal) shape can more closely be related to circles than the more quadratical shapes in case of the Reuter grid. (Even though it has to be notated that such a quadratical surface element is not valid for every point on a Reuter grid.) Fig. 3.13 reveals that not only the overall maximal distance is smaller for the triangle vertex grid, but that, for the majority of points, this is also true for the maximal distance within each individual surface element. This also supports the conclusion that the surface elements have a more circular shape. The triangle vertex grid exhibits a more isotropic structure, as the distance from one grid point to its adjacent grid points is approximately the same in all directions, which is not the case for the Reuter grid. (Adjacency is here defined for grid points with contacting Voronoi cells.)

Potential Energy of Charged Particles Another frequently-used definition of 'evenness' of a distribution is the potential energy function of charged particles located at the I grid points. This implies that, in a global sense, points are as 'far away' from each other as possible. The task of minimzing this potential energy is also known as 'Thomson's problem' (THOMSON 1904) and further described, for example, by EDMUNDSON (1992). The potential energy is defined as a sum of the reciprocal distances between two points. If the spherical distance is applied, then it holds

$$E = \sum_{i=0}^{I} \sum_{j=i+1}^{I-1} \frac{1}{\psi(\mathbf{x}_i, \mathbf{x}_j)}. \tag{3.121}$$

Again a normalization has to be applied to account for the different numbers of points for each grid. Since in Eq. (3.121) $I(I-1)/2$ terms are summed up, the result is divided by this factor,

$$\bar{E} = \frac{E}{I(I-1)/2}. \tag{3.122}$$

The results are listed in Tab. 3.4. In contrast to the minimum distance between grid points, as given by Tab. 3.2, here not only the nearest neighbors are regarded, but the reciprocal distance between each point and every other point is taken into account. Obviously, the potential energy is quite similar for all four point distributions. Nevertheless, it can again be observed that the triangle vertex and the Reuter grid exhibit the smallest potential energy, and the triangle center and quasi random grid are less evenly distributed with respect to the minimum energy criterion.

grid	I	normalized energy [1/rad]
Vertex22	5292	0.910989
Center4	5120	0.911538
Reuter64	5180	0.911071
Random5200	5200	0.911973

Table 3.4: Normalized potential energy of the different point distributions

3.5.2.2 Flexible Choice of Grid Points

The second requirement to be demanded of a nodal point pattern for the location of radial basis functions is the flexibility of the number of grid points. This is particularly important for the modeling of gravity field functionals, as the number of spline nodal points equals the number of unknown parameters, and thus it specifies the resolution of the solution. Among the discussed point distributions, the quasi random grid has the advantage that any arbitrary number of grid points can be applied. For the other point distributions, the restrictions in the choice of the number of points as given by Eqs. (3.110), (3.109), and (3.105), have to be considered. The grid triangle vertex is the least flexible, as the quantity of points depends exponentially on the chosen level of densification by a factor of 4^n. During each densification step, the triangles are divided into four smaller triangles. This does not leave many possibilities in the choice of the number of unknown parameters, as specified for different levels of densification in Tab. 3.5. The resolution is compared to the maximum degree of a corresponding spherical harmonic expansion with the same number of unknown parameters.

level	I	SH degree
3	1280	35
4	5120	70
5	20480	142
6	81920	285
7	327680	571

Table 3.5: Number of unknown parameters for different levels of densification of the grid triangle center and the corresponding spherical harmonic degree with a comparable number of coefficients

The grids triangle vertex and Reuter are adjustable much more easily to a required resolution, as their number of grid points increases quadratically with the choice of the control parameter. Here the Reuter grid with an (approximate) factor of $\frac{4}{\pi}$, instead of 10 in case of the grid triangle vertex, is even slightly more flexible.

3.5.2.3 Discussion

The four grids – triangle vertex, triangle center, Reuter grid, and the quasi random grid – were examined regarding the question which of them is most suitable to be used as nodal point pattern for the location of space localizing basis functions. The applied criteria were the evenness of the distribution (defined by the criterion of equal area, minimal distance between neighboring grid points, maximum distance between any point on the sphere to its nearest grid point, and minimum energy) on the one hand and the flexibility in the choice of the number of grid points on the other hand. Regarding the even distribution, it became apparent that the Reuter grid and the triangle vertex grid are superior to the other two options, as they perform better in all four tested criteria. The Reuter grid exhibits the most uniform area sizes for all points. The standard deviation of the area sizes is significantly smaller and the respective histogram supports this conclusion as well. Here the grid triangle vertex features the second smallest standard deviation, and the histogram shows that the area sizes are not as compact as in case of the Reuter grid, but still fairly close together. The area sizes associated with the triangle center grid and the pseudo random grid differ considerably more. Concerning the minimal distance between grid points, the triangle vertex and the Reuter grid are approximately equal, and again superior to the other two point distributions. Especially the pseudo random grid is out of the question, as the recursive pseudo random sequence produces points that can, in single cases, be located very close to each other. This can also be observed in Fig. 3.7. The distance between arbitrary points on the sphere and the nearest of the grid points is substantially the smallest in case of the grid triangle vertex. Here the surface elements have a shape that best resembles a spherical cap, which is reasonable for the arrangement of isotropic basis functions. Second best performs the Reuter grid, the two other point distributions are again less suitable. Concerning the minimum energy criterion, the four grid versions show approximately similar results, the triangle vertex grid being slightly ahead, followed by the Reuter grid. Regarding the flexibility in the number of grid points, the quasi random grid with its freely selectable quantity of points is advantageous, and the triangle center grid is unfavorable due to the exponential dependency on the level of densification. The Reuter grid and the triangle vertex grid are sufficiently flexible for the task of representing gravity field functionals of a given resolution, with the Reuter grid showing a slightly larger variety of possible grid point quantities. Concluding, it can be suggested that both, the Reuter grid and the triangle vertex grid, are very well suitable as nodal point pattern for space localizing basis functions, especially as they show a more uniform point pattern than the other two options. This result is confirmed by the visual impression when examining the displayed grids in Fig. 3.7 which presents the triangle vertex and Reuter grid as the most homogeneous point distributions.

From the listed criteria, the one specifying the maximum distance between arbitrary points and the nearest of the grid points is considered slightly more significant than the other ones. It accounts for the isotropic characteristics of the grid and thus indicates how well the surface elements resemble spherical caps. For the location of isotropic basis functions, this is considered more important than, for example, the exact sizes of the surface elements. Therefore, in the following considerations, the triangle vertex grid is chosen as point distribution for the arrangement of the spline functions.

3.5.3 Resolution

In order to compare field representations using different sets of basis functions, one inevitably has to deal with the definition of resolution. How can spatial resolution be compared when different types of basis function are applied? In the case relevant for this thesis, one has to deal with space localizing basis functions, defined on a grid and therefore given directly in the space domain on the one hand and spectral models such as spherical harmonics on the other hand. In case of space localizing basis functions, the resolution of a certain field representation is characterized by the number of basis functions per area. The number of field parameters matches the number of basis functions. And thus the spatial resolution can be regarded as the distance between nodal points on the grid of kernel functions. As also mentioned by LAPRISE (1992), there is no straightforward way to allocate a certain grid size to a given spectral model such as a spherical harmonic expansion. In this case the determination of an effective resolution is more complicated. Concerning the question of converting a spherical harmonic expansion to a spatial resolution, see also RUMMEL (1992). A quite common way to define the resolution of spherical harmonics is given by the size of half a wavelength of the shortest resolved zonal wave at the equator. A spherical harmonic expansion provides an isotropic and uniform resolution on the sphere, therefore the wavelength at the equator can be regarded as representative for the whole sphere. The shortest zonal wave corresponding to a certain maximum degree N is $|m| = N$. This leads to a measurement for the resolution of

$$r = \frac{\pi}{N}. \tag{3.123}$$

Another way of defining the resolution of a field represented by spherical harmonics could be proposed in a similar way as in case of space localizing basis functions. This means by defining the resolution by the area on the Earth's surface assigned to each unknown parameter. This leads to an area of

$$A = \frac{4\pi}{(N+1)^2} \tag{3.124}$$

for each coefficient with respect to the unit sphere. To convert this to a spatial resolution, different approaches could be thought of. Firstly, the aperture angle of a spherical cap with equal area can be calculated as follows,

$$\psi_1 = 2 \cdot \arccos\left(1 - \frac{A}{2\pi}\right). \tag{3.125}$$

Alternatively, the resolution can be defined as the square root of the area per unknown parameter (3.124) according to

$$\psi_2 = \sqrt{A}. \tag{3.126}$$

Furthermore, the number of unknown parameters can be distributed over the nodes of an evenly distributed grid of a comparable number of points, and the average distance between adjacent points can actually be calculated. This was performed for the triangle vertex grid (as this was chosen as nodal point pattern for the spline functions). The results of the different definitions of resolution are compared in Tab. 3.6. The first column of the table refers to the maximum degree of the spherical harmonic expansion. The second column specifies half of the wavelength of the shortest zonal wave at the equator. In the third column, the number of unknown parameters corresponding to N is given, which leads to an average area per unknown parameter given in column four according to Eq. (3.124). Resulting from this, the aperture angle of the corresponding spherical cap is listed in column five and the square root of the area element in column six. Columns seven and eight specify the level of the triangle vertex grid with comparable resolution and the resulting average distance between adjacent grid points. Obviously, the square root of the average surface area ψ_2 corresponds very well to the average distance calculated between the nodal points of the given grid. In contrast to this, taking half of the wavelength at the equator gives results slightly smaller, and the aperture angle of the spherical cap delivers larger values.

N	0.5 wavelength [°]	# unknowns	area per unknown	ψ_1[°]	ψ_2[°]	level	minDist [°]
60	3	3721	0.00338	3.75	3.33	19	3.30
120	1.5	14641	0.00086	1.89	1.68	38	1.69
200	0.9	40401	0.00031	1.14	1.01	63	1.03

Table 3.6: Spatial resolution of a spherical harmonic expansion

When a representation by space localizing basis functions is expected to deliver the same spatial resolution as a spherical harmonic model, in this thesis, the comparison is performed via the number of unknown parameters. This is based on the assumption that a uniformly distributed spline representation offers the same amount of information, enclosed in a certain number of unknown parameters, as a spherical harmonic model with its uniform and isotropic resolution on the sphere. Therefore, a desired resolution referring to space localizing basis functions can be obtained from the maximum degree N of a corresponding spherical harmonic expansion by calculating the number of spline kernels according to $I = (N+1)^2$. The respective resolution is then calculated from the average distance of the nodal points which agrees quite well with the square root of the average surface element per point.

4. Setting up the Observation Equations

The purpose of the satellite missions CHAMP, GRACE, and GOCE is to determine the gravity field of the Earth as accurately as possible. Therefore, a relationship has to be established between the (unknown) gravity field parameters and the provided observations. This is obtained via a functional model that links the particular observations to the gravity field parameters. Combining the observation equations for the different observation times leads to a (linearized) system of equations that can be solved by a least-squares adjustment procedure. In the first part of this chapter, it will shortly be reviewed how the unknown parameters can be estimated from a given set of observation equations in general. Subsequently, different gravity field functionals will be expressed in terms of a modeling by radial basis functions. This is a necessary tool to set up the functional models for a regional analysis. These models will then be described in the following section, leading to the observation equations for the different types of satellite observations. In detail, these types of observations can be characterized as satellite-to-satellite tracking in the high-low and in the low-low mode and by the concept of satellite gravity gradiometry.

4.1 Least Squares Approximation

In the following, it will be described how the unknown gravity field parameters are estimated from a given set of observations. This estimation process represents a standard Gauss-Markoff-Modell, as described, for example, by Koch (1997) or Niemeier (2002). All n observations can be arranged in a vector $\bar{\mathbf{y}}$ with the dimension $n \times 1$. The u unknown parameters can also be combined in a column vector $\bar{\mathbf{x}}$ with the dimension $u \times 1$. The observations can be linked to the unknown parameters via a functional model $f(\bar{\mathbf{x}})$ and can then be formulated in terms of this model and an additional measurement noise ϵ according to

$$\bar{\mathbf{y}} = f(\bar{\mathbf{x}}) + \epsilon. \tag{4.1}$$

Typically, the number of observations $n = \dim \bar{\mathbf{y}}$ is considerably larger than the number of unknown parameters $u = \dim \bar{\mathbf{x}}$. If the model is non-linear, a linearization becomes inevitable, therefore approximate values for the unknown parameters have to be introduced, and approximate observations can be calculated as functions of the approximate values. In case of a linear model, it is recommended to start with approximate values for the unknown parameters as well and to calculate their influence on the observations,

$$\mathbf{y}_0 = f(\mathbf{x}_0). \tag{4.2}$$

The linearization procedure of the model f with respect to the unknown parameters can be performed by a Taylor expansion truncated after the linear term,

$$\bar{\mathbf{y}} = \mathbf{y}_0 + \left.\frac{\partial f(\bar{\mathbf{x}})}{\partial \bar{\mathbf{x}}}\right|_0 (\bar{\mathbf{x}} - \mathbf{x}_0) + ... \tag{4.3}$$

The reduced observations and the corrections to the unknown parameters are then calculated according to

$$\mathbf{y} = \bar{\mathbf{y}} - \mathbf{y}_0 \qquad \text{and} \qquad \mathbf{x} = \bar{\mathbf{x}} - \mathbf{x}_0. \tag{4.4}$$

The partial derivatives of the linearization (4.3) can be combined in the design matrix $\mathbf{A}$. It is of dimension $n \times u$, and its elements are defined by the partial derivatives of the function $f_k(\mathbf{x})$ with respect to x_i,

$$(A)_{ki} = \left.\frac{\partial f_k(\bar{\mathbf{x}})}{\partial x_i}\right|_0. \tag{4.5}$$

This leads to the linear system of equations

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon \qquad \text{with} \qquad \mathcal{C}(\epsilon) = \sigma^2 \mathbf{P}_\epsilon^{-1}. \tag{4.6}$$

The standard Gauss-Markov model is based on the assumptions that the measurement errors have an expectation of zero and that the covariance matrix of the observations is known a-priori,

$$E\{\epsilon\} = \mathbf{0} \qquad \text{and} \qquad \mathcal{C}(\epsilon) = \mathcal{C}(\mathbf{y}) = \sigma^2 \mathbf{P}_\epsilon^{-1}. \tag{4.7}$$

Here σ stands for the unknown variance factor, and $\mathbf{P}_\epsilon$ denotes the weight matrix of the observations. An estimation by least squares adjustment corresponds to the minimization of the square sum of the residuals,

$$\Omega = \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{A}\mathbf{x})^T \mathbf{P}_\epsilon (\mathbf{y} - \mathbf{A}\mathbf{x}). \tag{4.8}$$

The minimum condition can be obtained by differentiation according to

$$\frac{\partial \Omega}{\partial \hat{\mathbf{x}}} = 2\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A}\mathbf{x} - 2\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y} = 0. \tag{4.9}$$

This leads to the following system of normal equations,

$$\mathbf{N}\mathbf{x} = \mathbf{n} \qquad \text{with} \quad \mathbf{N} = \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A}\mathbf{x} \quad \text{and} \quad \mathbf{n} = \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y}. \tag{4.10}$$

The solution of the normal equations yields the estimation of the unknown parameters by means of least squares adjustment,

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y} = \mathbf{N}^{-1}\mathbf{n}. \tag{4.11}$$

This solution corresponds to the best linear unbiased estimate and results in the maximization of the likelihood function (Koch 1997). The covariance matrix of the unknown parameters $\mathcal{C}(\hat{\mathbf{x}})$ can be derived by applying the law of covariance propagation to (4.11) under consideration of (4.6),

$$\mathcal{C}(\hat{\mathbf{x}}) = \sigma^2 \mathbf{N}^{-1}. \tag{4.12}$$

The unknown variance factor can be estimated by

$$\hat{\sigma} = \frac{1}{n-u}(\mathbf{y} - \mathbf{A}\hat{\mathbf{x}})^T \mathbf{P}_\epsilon (\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}), \tag{4.13}$$

which leads to the estimated covariance matrix of the unknown parameters,

$$\hat{\mathcal{C}}(\hat{\mathbf{x}}) = \hat{\sigma}^2 \mathbf{N}^{-1}. \tag{4.14}$$

4.2 The Gravity Field and its Functionals in Terms of Splines

In this section, the basis functions, as introduced in Section 3.4.2, will be adopted according to the specific task of parameterization and determination of the gravity field with its different functionals. The mathematical considerations above have primarily been performed for functions given on the surface of the (unit) sphere. When dealing with satellite data, on the contrary, the observation points are located in the exterior of the Earth. Therefore, the outward continued basis functions (3.75) have to be applied, and a reference sphere Ω_R with radius R approximating the semi-major axis of the Earth is introduced. The spline kernels are located at nodal points $\mathbf{r}_i$ on Ω_R with

$$|\mathbf{r}_i| = R. \tag{4.15}$$

Arbitrary points in the exterior of this sphere, for example at the satellite's positions, are denoted by $\mathbf{r}$ with

$$|\mathbf{r}| = r. \tag{4.16}$$

For simplification reasons, the spline kernel located at a specific nodal point $\mathbf{r}_i$ will be denoted as $\Phi_i(\mathbf{r})$ in the following according to

$$\Phi_i(\mathbf{r}) := \Phi(\mathbf{r}, \mathbf{r}_i). \tag{4.17}$$

The spline kernel evaluated at the point $\mathbf{r}$ reads

$$\Phi_i(\mathbf{r}) = \frac{GM}{R} \sum_{n=2}^{\infty} k_n \left(\frac{R}{r}\right)^{n+1} P_n(t), \tag{4.18}$$

with

$$t = \cos(\psi) = \frac{\mathbf{r}}{r} \cdot \frac{\mathbf{r}_i}{R}. \tag{4.19}$$

The factor GM/R is introduced into the basis functions, so that the spline coefficients a_i in (3.74) become dimensionless values. In the following sections, the representation of the different gravity field functionals in terms of basis functions defined by (4.18) will be specified. These gravity field functionals are the gravitational potential, gravity, and the gravity gradient.

4.2.1 Gravitational Potential

The gravitational potential in terms of spline functions (4.18) can be represented as follows,

$$V(\mathbf{r}) = \sum_{i=1}^{I} a_i \Phi_i(\mathbf{r}). \tag{4.20}$$

The potential as well as gravity and the gravity gradient are linear functionals of the unknown parameters a_i, thus they can be phrased as a matrix-vector product. Therefore, the unknown parameters are arranged in the vector $\mathbf{x}$,

$$\mathbf{x} = (a_0, ..., a_I)^T, \tag{4.21}$$

and the gravitational potential at N positions of the satellite's orbit can then be expressed by

$$\begin{pmatrix} V(\mathbf{r}_1) \\ \vdots \\ V(\mathbf{r}_N) \end{pmatrix} = \mathbf{V}\mathbf{x}, \tag{4.22}$$

with the matrix $\mathbf{V}$ consisting of the basis functions evaluated at the satellite's positions $\mathbf{r}_k$ with $k = 1, ..., N$,

$$\mathbf{V} = \begin{pmatrix} \Phi_1(\mathbf{r}_1) & \Phi_2(\mathbf{r}_1) & \dots & \Phi_I(\mathbf{r}_1) \\ \Phi_1(\mathbf{r}_2) & \Phi_2(\mathbf{r}_2) & \dots & \Phi_I(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \Phi_1(\mathbf{r}_N) & \Phi_2(\mathbf{r}_N) & \dots & \Phi_I(\mathbf{r}_N) \end{pmatrix}_{N \times I} . \tag{4.23}$$

4.2.2 Gravity

Under consideration of Eq. (3.5), gravity at positions along the arcs of the satellite's orbit can be calculated according to

$$\mathbf{g}(\mathbf{r}) = \nabla V(\mathbf{r}) = \sum_{i=1}^{I} a_i \nabla \Phi_i(\mathbf{r}). \tag{4.24}$$

Again, this can be expressed by the corresponding matrix-vector product

$$\begin{pmatrix} \mathbf{g}(\mathbf{r}_1) \\ \vdots \\ \mathbf{g}(\mathbf{r}_N) \end{pmatrix} = \bar{\mathbf{G}}\mathbf{x}, \tag{4.25}$$

where the elements of the matrix $\bar{\mathbf{G}}$ contain the gradients of the basis functions,

$$\bar{\mathbf{G}} = \begin{pmatrix} \nabla\Phi_1(\mathbf{r}_1) & \nabla\Phi_2(\mathbf{r}_1) & \dots & \nabla\Phi_I(\mathbf{r}_1) \\ \nabla\Phi_1(\mathbf{r}_2) & \nabla\Phi_2(\mathbf{r}_2) & \dots & \nabla\Phi_I(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \nabla\Phi_1(\mathbf{r}_N) & \nabla\Phi_2(\mathbf{r}_N) & \dots & \nabla\Phi_I(\mathbf{r}_N) \end{pmatrix}_{3N\times I} . \tag{4.26}$$

Usually, the gravity vector is obtained by differentiating the gravitational potential with respect to the coordinates of a local north-oriented frame and a subsequent rotation into a cartesian Earth-fixed reference frame. Here a different method is applied, as the derivatives of the basis functions are directly calculated with respect to the cartesian coordinates of an Earth-fixed frame by applying the chain rule,

$$\nabla\Phi = \begin{pmatrix} \partial\Phi/\partial x \\ \partial\Phi/\partial y \\ \partial\Phi/\partial z \end{pmatrix} = \begin{pmatrix} \partial\Phi/\partial r \cdot \partial r/\partial x + \partial\Phi/\partial t \cdot \partial t/\partial x \\ \partial\Phi/\partial r \cdot \partial r/\partial y + \partial\Phi/\partial t \cdot \partial t/\partial y \\ \partial\Phi/\partial r \cdot \partial r/\partial z + \partial\Phi/\partial t \cdot \partial t/\partial z \end{pmatrix} , \tag{4.27}$$

with the partial derivatives of the basis functions with respect to r and t,

$$\frac{\partial\Phi}{\partial r} = \sum_{n=2}^{\infty} \frac{-(n+1)}{R} k_n \left(\frac{R}{r}\right)^{n+2} P_n(t), \tag{4.28}$$

$$\frac{\partial\Phi}{\partial t} = \sum_{n=2}^{\infty} k_n \left(\frac{R}{r}\right)^{n+2} \frac{dP_n(t)}{dt}, \tag{4.29}$$

and with the partial derivatives of r and t with respect to the cartesian coordinates,

$$\frac{\partial r}{\partial\alpha} = \frac{\alpha}{r} \qquad \text{and} \qquad \frac{\partial t}{\partial\alpha} = \frac{\alpha_i}{rR} - \frac{t\,\alpha}{r^2} \qquad \text{with} \qquad \alpha = x, y, z. \tag{4.30}$$

Here α represents the coordinates of $\mathbf{r}$ and α_i the coordinates of $\mathbf{r}_i$. The positions used for the calculation of the matrix in Eq. (4.26) have to be given in an Earth-fixed co-rotating coordinate system, thus gravity in Eq. (4.25) is given in this Earth-fixed frame as well. In contrast to this, the functionals of the satellite movement that will serve as functional models in the gravity field determination process, as will be described in Section 4.3, refer to the inertial reference frame. Therefore, they require gravity to be formulated in this frame. The corresponding transformation can be expressed as

$$\mathbf{G} := \mathbf{R}\bar{\mathbf{G}}, \tag{4.31}$$

with $\mathbf{G}$ denoting the matrix related to the inertial reference frame. Using the rotation matrices of the Earth rotation, as given in the IERS Conventions (McCarthy and Petit 2004) of the International Earth Rotation and Reference System Service (IERS), the relationship between the International Celestial Reference Frame (ICRF) as a realization of the quasi-inertial system and the International Terrestrial Reference Frame (ITRF) as realization of an Earth fixed reference system can be established. The rotations from the ITRF to the ICRF for each observation time can be given by the matrices $\mathbf{R}(t_k)$ arranged in the blockdiagonal matrix

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}(t_1) & & \\ & \ddots & \\ & & \mathbf{R}(t_N) \end{pmatrix} . \tag{4.32}$$

Gravity in the ICRF can be obtained by inserting $\mathbf{G}$ instead of $\bar{\mathbf{G}}$ in Eq. (4.25).

4.2.3 Gravity Gradient

In case of satellite gravity gradiometry, the gravity gradient components serve as observations. Therefore, the functional model for this type of observations is required as well. The gravity gradient is calculated as second derivatives of the gravitational potential, expressed in terms of spline functions according to

$$\nabla\nabla V(\mathbf{r}) = \sum_{i=1}^{I} a_i \nabla\nabla\Phi_i(\mathbf{r}). \tag{4.33}$$

The corresponding matrix-vector product can be formulated as follows,

$$\begin{pmatrix} \nabla\nabla V(\mathbf{r}_1) \\ \vdots \\ \nabla\nabla V(\mathbf{r}_N) \end{pmatrix} = \bar{\mathbf{T}}\mathbf{x}, \tag{4.34}$$

with the block matrix $\bar{\mathbf{T}}$ containing the second derivatives of the basis functions,

$$\bar{\mathbf{T}} = \begin{pmatrix} \nabla\nabla\Phi_1(\mathbf{r}_1) & \nabla\nabla\Phi_2(\mathbf{r}_1) & \dots & \nabla\nabla\Phi_I(\mathbf{r}_1) \\ \nabla\nabla\Phi_1(\mathbf{r}_2) & \nabla\nabla\Phi_2(\mathbf{r}_2) & \dots & \nabla\nabla\Phi_I(\mathbf{r}_2) \\ \vdots & \vdots & & \vdots \\ \nabla\nabla\Phi_1(\mathbf{r}_N) & \nabla\nabla\Phi_2(\mathbf{r}_N) & \dots & \nabla\nabla\Phi_I(\mathbf{r}_N) \end{pmatrix}. \tag{4.35}$$

Here $\nabla\nabla\Phi$ denotes the Hessian matrix of the second derivatives of a basis function,

$$\nabla\nabla\Phi = \left\{ \frac{\partial^2\Phi}{\partial\alpha\partial\beta} \right\}_{\alpha\beta} \qquad \text{with} \qquad \alpha, \beta = x, y, z. \tag{4.36}$$

Depending on the specific application, these second derivatives can be arranged either in a 3 × 3 tensor or in a column vector, as will be shown below. Again the chain rule is applied in order to directly calculate the derivatives in the cartesian Earth-fixed frame,

$$\begin{aligned} \Phi_{\alpha\beta} = \Phi_{\beta\alpha} = & \frac{\partial\Phi}{\partial r} \cdot \frac{\partial^2 r}{\partial\alpha\partial\beta} + \frac{\partial\Phi}{\partial t} \cdot \frac{\partial^2 t}{\partial\alpha\partial\beta} + \frac{\partial^2\Phi}{\partial r^2} \cdot \frac{\partial r}{\partial\alpha}\frac{\partial r}{\partial\beta} \\ & + \frac{\partial^2\Phi}{\partial r\partial t} \cdot \frac{\partial r}{\partial\alpha}\frac{\partial t}{\partial\beta} + \frac{\partial^2\Phi}{\partial r\partial t} \cdot \frac{\partial r}{\partial\beta}\frac{\partial t}{\partial\alpha} + \frac{\partial^2\Phi}{\partial t^2}\frac{\partial t}{\partial\alpha}\frac{\partial t}{\partial\beta}, \end{aligned} \tag{4.37}$$

with the second derivatives of the basis functions with respect to r and t,

$$\frac{\partial^2\Phi}{\partial r^2} = \sum_{n=2}^{\infty} \frac{(n+2)(n+1)}{R^2} k_n \left(\frac{R}{r}\right)^{n+3} P_n(t), \tag{4.38}$$

$$\frac{\partial^2\Phi}{\partial r\partial t} = \sum_{n=2}^{\infty} \frac{(n+1)}{R} k_n \left(\frac{R}{r}\right)^{n+2} \frac{dP_n(t)}{dt}, \tag{4.39}$$

$$\frac{\partial^2\Phi}{\partial t^2} = \sum_{n=2}^{\infty} k_n \left(\frac{R}{r}\right)^{n+2} \frac{d^2P_n(t)}{dt^2}. \tag{4.40}$$

The second derivatives of r and t with respect to the cartesian coordinates can be formulated as follows,

$$\frac{\partial^2 r}{\partial\alpha\partial\beta} = \frac{1}{r}\delta_{\alpha\beta} - \frac{\alpha\beta}{r^3} \qquad \text{and} \qquad \frac{\partial^2 t}{\partial\alpha\partial\beta} = -\frac{t}{r^2}\delta_{\alpha\beta} - \frac{\alpha_i\beta + \alpha\beta_i}{r^3 R} + \frac{3\alpha\beta\, t}{r^4} \qquad \text{with} \qquad \alpha, \beta = x, y, z. \tag{4.41}$$

Again α and β are the coordinates of $\mathbf{r}$ and α_i and β_i the coordinates of $\mathbf{r}_i$. The observations of the second derivatives are given in the reference frame of the gradiometer, the so-called 'gradiometer reference frame' (GRF), orientated along the three axes of the gradiometer instrument with origin in the nominal intersection of the three one-axis gradiometers. In contrast to that, the derivatives in the formulas given above are formulated in an Earth-fixed reference frame (ITRF) with origin located in the geocenter, z-axis directed to the pole, x-axis fixed in the equatorial plane in the direction of the Greenwich meridian, and y-axis as completion to a right-handed system. Therefore, the observation equation has to be transformed from the ITRF into the GRF. When the second derivatives for each observation point $\mathbf{r}_k$ are arranged according to

$$\bar{\mathbf{T}}(\mathbf{r}_k) = \nabla\nabla\Phi = \begin{pmatrix} \Phi_{xx} & \Phi_{xy} & \Phi_{xz} \\ \Phi_{yx} & \Phi_{yy} & \Phi_{yz} \\ \Phi_{zx} & \Phi_{zy} & \Phi_{zz} \end{pmatrix}, \tag{4.42}$$

this transformation can be performed by using the transformation formula for tensors,

$$\mathbf{T}(\mathbf{r}_k) = \mathbf{D}(\mathbf{r}_k)\bar{\mathbf{T}}(\mathbf{r}_k)\mathbf{D}(\mathbf{r}_k)^T. \tag{4.43}$$

$\bar{\mathbf{T}}$ is the tensor of second derivatives in the Earth-fixed frame and $\mathbf{T}$ the tensor referring to the gradiometer reference frame. The matrix $\mathbf{D}$ is the 3×3 rotation matrix transforming the former frame into the latter. Further information about the transformation of gravity gradients from one coordinate system into another one can be found in Koop (1993). For practical calculations, it can often be reasonable to combine the independent tensor components in a column vector according to

$$\bar{\mathbf{t}}(\mathbf{r}_k) = (\Phi_{xx}, \Phi_{xy}, \Phi_{xz}, \Phi_{yy}, \Phi_{yz}, \Phi_{zz})^T. \tag{4.44}$$

Then the transformation described by Eq. (4.43) can be reformulated as a single matrix-vector operation, as, for example, described by Ditmar et al. (2003),

$$\mathbf{t}(\mathbf{r}_k) = \mathbf{R}(\mathbf{r}_k)\bar{\mathbf{t}}(\mathbf{r}_k), \tag{4.45}$$

with $\mathbf{t}$ denoting the rotated values. The matrix $\mathbf{R}(\mathbf{r}_k)$ relates the vector $\mathbf{t}(\mathbf{r}_k)$ to the vector $\bar{\mathbf{t}}(\mathbf{r}_k)$; it emerges from carrying out the two rotations in Eq. (4.43). When the transformation matrices for the different observation points are combined into one blockdiagonal matrix according to

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}(\mathbf{r}_1) & & \\ & \ddots & \\ & & \mathbf{R}(\mathbf{r}_N) \end{pmatrix}, \tag{4.46}$$

then the transformation can be applied to the matrix $\bar{\mathbf{T}}$ in Eq. (4.35) as follows,

$$\mathbf{T} = \mathbf{R}\bar{\mathbf{T}}. \tag{4.47}$$

The gravity gradient related to the GRF can be obtained by inserting $\mathbf{T}$ instead of $\bar{\mathbf{T}}$ in Eq. (4.34).

4.3 Functional Model

In the following section, the functional models providing the relationship between the unknown gravity field parameters and the different types of observations provided by the satellite missions CHAMP, GRACE, and GOCE will be derived. These observation principles are the concept of precise orbit determination (POD) derived from satellite-to-satellite tracking in the high-low mode, satellite-to-satellite tracking in the low-low mode, and satellite gravity gradiometry. The functional models can be applied to different gravity field representations, e.g. spherical harmonic expansions or radial basis functions. The representation of the different gravity field functionals in terms of radial basis functions, as described in Section 4.2, leads to the specific observation equations for the regional gravity field recovery. To exploit all the advantages of a regional gravity field recovery, only the data being associated with the respective regional area has to be used. Therefore, the observation equations are established for satellite data over the selected regional recovery area, while the coverage with satellite data should be slightly larger than the recovery region itself to prevent the solution from geographical truncation effects. This aspect has to be considered with care especially for the long wavelength gravity field features. Thus the orbit information and additional satellite data have to be cut out over the regional area. Consequently, the employed observation models necessarily have to be based on short arcs of the satellite's orbit. The use of short arcs presents further advantages. Unmodeled disturbing forces acting on the satellite do not accumulate and, therefore, only have a limited impact on the solution. A further benefit of short arcs lies in the fact that one can deal with data gaps more easily, as a new arc can be started after each data gap. In this way, the observations along an arc can be regarded to have equal distances in time without gaps. Therefore, there is no necessity to consider respective exceptions during the data processing.

4.3.1 Precise Orbit Determination

All three satellite missions CHAMP, GRACE, and GOCE carry a receiver for the Global Positioning System (GPS) in order to acquire satellite-to-satellite tracking data in the high-low configuration. From this data, the satellites' orbits can be calculated by means of precise orbit determination (POD), see, for example, Švehla and Rothacher (2001). In case of CHAMP, this POD information provides the primary type of observations, but it establishes the basis for the functional model of the mission GRACE as well and is also used in the GOCE processing. Different approaches to process POD data have been proposed, among them are the acceleration approach, as developed by Austen and Reubelt (2000), the acceleration approach using double differences, as introduced by Ditmar and van Eck van der Sluijs (2004), and the energy integral approach as a very popular tool in satellite geodesy. In the context of determining a gravity field solution from CHAMP data, it was, for example, used by Gerlach et al. (2003). An overview of the different methods is given in Ilk et al. (2005a). In the following, only the approach actually used for the calculation procedure in this thesis will be described. It is based on an integral equation approach, first proposed by Schneider (1968) in the context of orbit determination and then applied to the gravity field determination by Reigber (1969). The following procedure was applied successfully to the calculation of a global CHAMP gravity field model in the space domain by Mayer-Gürr et al. (2005). More details of the method can be found in Mayer-Gürr (2006).

The relation between the satellite's movement along the orbit and the forces $\mathbf{f}$ acting on the satellite is described by Newton's equation of motion, here formulated with respect to a unity mass element,

$$\ddot{\mathbf{r}}(t) = \mathbf{f}(t; \mathbf{r}, \dot{\mathbf{r}}), \tag{4.48}$$

with $\ddot{\mathbf{r}}(t)$ denoting the satellite's acceleration for a given time t and $\mathbf{r}$ and $\dot{\mathbf{r}}$ describing the satellite's position and velocity, respectively. Typically, this differential equation is non-linear, as the force function does not depend linearly on the satellite's positions. Integrating the differential equation twice results in an integral equation for the satellite's orbit. With the boundary values

$$\mathbf{r}_A := \mathbf{r}(t_A), \quad \mathbf{r}_B := \mathbf{r}(t_B), \quad t_A < t_B, \tag{4.49}$$

for the start and end position of the arc of the satellite's orbit. This integral equation can be formulated as the solution of a boundary value problem according to

$$\mathbf{r}(\tau) = (1-\tau)\,\mathbf{r}_A + \tau \mathbf{r}_B - T^2 \int\limits_{\tau'=0}^{1} K(\tau, \tau')\, \mathbf{f}(\tau', \mathbf{r}, \dot{\mathbf{r}})\, d\tau', \tag{4.50}$$

with the normalized time

$$\tau = \frac{t - t_A}{T} \qquad \text{with} \qquad T = t_B - t_A \tag{4.51}$$

and the integral kernel

$$K(\tau, \tau') = \begin{cases} \tau\,(1-\tau'), & \tau \le \tau', \\ \tau'\,(1-\tau), & \tau' \le \tau, \end{cases} . \tag{4.52}$$

Eq. (4.50) is a Fredholm type integral equation of the second kind. The unknowns in this equation are the boundary values $\mathbf{r}_A$ and $\mathbf{r}_B$ and the specific force function $\mathbf{f}(\tau', \mathbf{r}, \dot{\mathbf{r}})$ along the orbit. In the following, the positions $\mathbf{r}$ in the force function will be substituted by the positions $\mathbf{r}_\epsilon$ measured by GPS. In case of POD measurements, they can be assumed as sufficiently accurate, as has been investigated in Mayer-Gürr (2006). Therefore, the dependence of the force function on the unknown satellite orbit can be omitted. When the same integral equation is used to derive observation equations for the highly accurate satellite-to-satellite tracking measurements, this dependence has to be considered, as will be described in Section 4.3.1.1. The satellite's velocity $\dot{\mathbf{r}}$ only influences the force function due to the surface forces acting on the satellite. Those surface forces can directly be determined by the onboard accelerometer, therefore the dependence of the force function on the satellite's velocity can be neglected as well. This leads for the POD observations to the simplification $\mathbf{f}(\tau', \mathbf{r}, \dot{\mathbf{r}}) = \mathbf{f}(\tau')$.

As observations serve the satellite positions $\mathbf{r}(\tau_k)$ which are determined at N discrete, equidistant normed points in time τ_k calculated by

$$\tau_k = \frac{k-1}{N-1} \qquad \text{for } k = 1, ..., N. \tag{4.53}$$

As equation (4.50) is evaluated for each observed position $\mathbf{r}(\tau_k)$ along the arc of the satellite's orbit, one obtains a system of equations which can be formulated in matrix notation as follows,

$$\mathbf{l} = \mathbf{B}\mathbf{b} + \mathbf{h}. \tag{4.54}$$

The vector $\mathbf{l}$ represents the observed positions of the satellite,

$$\mathbf{l} = \begin{pmatrix} \mathbf{r}(\tau_1) \\ \vdots \\ \mathbf{r}(\tau_N) \end{pmatrix}. \tag{4.55}$$

The vector $\mathbf{b}$ contains the boundary values and the matrix $\mathbf{B}$ the corresponding normalized observation times,

$$\mathbf{B} = \begin{pmatrix} (1-\tau_1) & \tau_1 \\ (1-\tau_2) & \tau_2 \\ \vdots & \vdots \\ (1-\tau_N) & \tau_N \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \mathbf{r}_A \\ \mathbf{r}_B \end{pmatrix}. \tag{4.56}$$

The vector $\mathbf{h}$ stands for the integral evaluated at the N observation epochs,

$$\mathbf{h} = \begin{pmatrix} \mathbf{h}(\tau_1) \\ \vdots \\ \mathbf{h}(\tau_N) \end{pmatrix} \qquad \text{with} \qquad \mathbf{h}(\tau_k) = -T^2 \int_0^1 K(\tau_k, \tau')\mathbf{f}(\tau')\, d\tau'. \tag{4.57}$$

In order to establish a relation between the observations and the unknown parameters $\mathbf{x}$, the integral is expressed as a linearized functional of these parameters,

$$\mathbf{h} = \mathbf{h}_0 + \mathbf{A}\mathbf{x} \qquad \text{with} \qquad (\mathbf{A})_{ki} = \frac{\partial \mathbf{h}(\tau_k)}{\partial x_i}. \tag{4.58}$$

This results in the following linear system of observation equations,

$$\mathbf{l} - \mathbf{h}_0 = \mathbf{B}\mathbf{b} + \mathbf{A}\mathbf{x}, \tag{4.59}$$

with the vector of unknown parameters according to Eq. (4.21). The elements of matrix $\mathbf{A}$ can be derived by applying the chain rule, i.e. by partially differentiating the integral with respect to the force function and then differentiating $\mathbf{f}$ with respect to the unknown parameters,

$$(\mathbf{A})_{ki} = \frac{\partial \mathbf{h}(\tau_k)}{\partial x_i} = \frac{\partial \mathbf{h}(\tau_k)}{\partial \mathbf{f}(\tau')} \frac{\partial \mathbf{f}(\tau')}{\partial x_i}. \tag{4.60}$$

It reads expressed in matrix notation

$$\mathbf{A} = \mathbf{K}\,\mathbf{G} \qquad \text{with} \qquad (\mathbf{K})_{kj} = \frac{\partial \mathbf{h}(\tau_k)}{\partial \mathbf{f}_j} \qquad \text{and} \qquad (\mathbf{G})_{ji} = \frac{\partial \mathbf{f}_j}{\partial x_i}. \tag{4.61}$$

The set up of matrix **G** has already been described in Section 4.2.2; it is given according to Eq. (4.26) under consideration of Eq. (4.31). The specific forces along the satellite's orbit can then be expressed in terms of the unknown gravity field functionals according to

$$\begin{pmatrix} \mathbf{f}(\tau_1) \\ \vdots \\ \mathbf{f}(\tau_N) \end{pmatrix} = \mathbf{G}\mathbf{x} + \mathbf{a}, \tag{4.62}$$

with the vector **a** representing the reference gravity field, the tide models, and the disturbing surface forces as determined by the onboard accelerometer. The observation vector is reduced by these forces which implies that the accelerometer measurements and force models are assumed to be error-free. The matrix **K** represents the numerical integration of the integral (4.57). By means of quadrature methods, a linear relationship between **h** and **f** can be established,

$$\begin{pmatrix} \mathbf{h}(\tau_1) \\ \vdots \\ \mathbf{h}(\tau_N) \end{pmatrix} = \mathbf{K} \begin{pmatrix} \mathbf{f}(\tau_1) \\ \vdots \\ \mathbf{f}(\tau_N) \end{pmatrix}. \tag{4.63}$$

For more details concerning the calculation of **K**, refer to MAYER-GÜRR (2006). Using this matrix **K**, the approximate values $\mathbf{h}_0$ for the integral can be calculated from the reference accelerations according to

$$\mathbf{h}_0 = \mathbf{K}\mathbf{a}. \tag{4.64}$$

This leads to the following Gauss-Markoff model,

$$\mathbf{l} - \mathbf{K}\mathbf{a} = \mathbf{K}\mathbf{G}\mathbf{x} + \mathbf{B}\mathbf{b} + \epsilon, \tag{4.65}$$

where **b** and **x** are the vectors of unknown parameters, and ϵ denotes the noise.

4.3.1.1 Refinement of the Functional Model

The functional model described so far can be used in the processing of POD measurements. The observation equations can be modified and subsequently applied to establish the linearized model for intersatellite K-band measurements as in case of the GRACE mission as well. The superior accuracy of the K-band observations results in high demands on the accuracy of the functional model. The intersatellite ranges can be observed with an accuracy of a few μm, thus the errors occurring from the approximations made in the model itself have to be significantly smaller than these measurement errors. Those approximation errors can originate from the linearization procedure, from the numerical integration, and from values introduced into the functional model. The first one can be considered by good approximate values or by an iteration of the estimation procedure. The numerical integration is non-critical as well, as shown by MAYER-GÜRR (2006). But the accuracy of the satellite's positions that are introduced into the functional model poses a more serious problem. In the integral (4.57), the force function **f** is evaluated at the specific positions of the satellite at specific observation epochs. This results in the force function to depend not only on time, but on the position as well: $\mathbf{f}(\mathbf{r}(\tau))$. In case of the POD observations, the inaccuracies of the GPS positions were neglected, but this is not possible in case of K-band measurements. The positions can be determined with an accuracy of a few cm, which is not sufficient compared to the high precision measurements provided by the K-band instrument. Therefore, MAYER-GÜRR (2006) proposed the use of refined observation equations. This refinement procedure is shortly described in the following.

The evaluation of the integral at noisy positions $\mathbf{r}_\epsilon$ results in positions $\hat{\mathbf{r}}$,

$$\hat{\mathbf{r}}(\tau) = (1-\tau)\mathbf{r}_A + \tau\mathbf{r}_B + T^2 \int_0^1 K(\tau,\tau')\mathbf{f}(\mathbf{r}_\epsilon(\tau'))\, d\tau'. \tag{4.66}$$

The difference compared to the true positions $\mathbf{r}$ can be calculated according to

$$\mathbf{r}(\tau) - \hat{\mathbf{r}}(\tau) = T^2 \int_0^1 K(\tau, \tau') \left[\mathbf{f}(\mathbf{r}(\tau')) - \mathbf{f}(\mathbf{r}_\epsilon(\tau'))\right] d\tau'. \tag{4.67}$$

The above equation can be given in a simplified version,

$$\mathbf{r} - \hat{\mathbf{r}} = \mathcal{K}\left[\mathbf{f}(\mathbf{r}) - \mathbf{f}(\mathbf{r}_\epsilon)\right], \tag{4.68}$$

with the linear integral operator $\mathcal{K} = T^2 \int_0^1 K(\tau, \tau')(\cdot)\, d\tau'$. Linearizing the force function,

$$\mathbf{f}(\mathbf{r}) = \mathbf{f}(\mathbf{r}_\epsilon) + \nabla\mathbf{f}\big|_{\mathbf{r}_\epsilon} \cdot (\mathbf{r} - \mathbf{r}_\epsilon) + \ldots, \tag{4.69}$$

yields

$$\mathbf{r} - \hat{\mathbf{r}} = \mathcal{K}\, \nabla\mathbf{f} \cdot (\mathbf{r} - \mathbf{r}_\epsilon). \tag{4.70}$$

The insertion of Eq. (4.70) instead of $\hat{\mathbf{r}}$ into Eq. (4.66) and some rearrangements lead to

$$\left[\mathcal{I} - \mathcal{K}\, \nabla\mathbf{f}(\mathbf{r}_\epsilon)\right](\mathbf{r} - \mathbf{r}_\epsilon) = \mathcal{K}\, \mathbf{f}(\mathbf{r}_\epsilon) + \mathbf{b} - \mathbf{r}_\epsilon, \tag{4.71}$$

with $\mathcal{I}$ denoting the unity operator and $\mathbf{b}(\tau) = (1 - \tau)\mathbf{r}_A + \tau\mathbf{r}_B$ being the functional of the line of sight connection. As long as the inverse of the operator $[\mathcal{I} - \mathcal{K}\, \nabla\mathbf{f}(\mathbf{r}_\epsilon)]$ exists, the noise-free positions can be calculated during the linearization process of the force function,

$$\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}_\epsilon = \left[\mathcal{I} - \mathcal{K}\, \nabla\mathbf{f}(\mathbf{r}_\epsilon)\right]^{-1} \left[\mathcal{K}\, \mathbf{f}(\mathbf{r}_\epsilon) + \mathbf{b} - \mathbf{r}_\epsilon\right]. \tag{4.72}$$

This equation can be discretized analogously to the procedure used with the POD observations,

$$\Delta\mathbf{r} = \left(\mathbf{I} - \mathbf{K}\,\mathbf{T}\right)^{-1} \left(\mathbf{K}\,\mathbf{f} + \mathbf{B}\,\mathbf{b} - \mathbf{r}_\epsilon\right), \tag{4.73}$$

with $\mathbf{K}$ being the matrix of the numerical quadrature and $\mathbf{B}$ being the matrix containing the functional of the boundary values as in (4.56). $\mathbf{T}$ is the matrix of the gravity gradients according to

$$\mathbf{T} = \begin{pmatrix} \nabla\mathbf{f}(\tau_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \nabla\mathbf{f}(\tau_N) \end{pmatrix}. \tag{4.74}$$

Eq. (4.73) can now be used to calculate error-free positions,

$$\mathbf{r} = \mathbf{r}_\epsilon + \Delta\mathbf{r}. \tag{4.75}$$

4.3.2 Low-Low Satellite-to-Satellite Tracking

For the processing of satellite-to-satellite tracking data in the low-low configuration, as provided by the K-band measurement instrument onboard the GRACE mission, different approaches have been proposed as well. Among them are the employment of the energy integral, as proposed by Jekeli (1999), the acceleration approach, as, for example, used by Sharifi and Keller (2005), and the Hammerstein-Schneider method, see, for example, Ilk (1984). The approach that is used in the calculations presented here is based on the integral equation described for the POD observations in Eq. (4.50). This method is explained in more detail in Mayer-Gürr et al. (2006); in the following it will shortly be reviewed.

If precise intersatellite functionals as line-of-sight ranges or range-rate measurements are available as in case of the GRACE mission, the mathematical model can be derived by projecting the relative vector between the two satellites onto the line-of-sight connection,

$$\rho(\tau) = \mathbf{e}_{12}(\tau) \cdot \left(\mathbf{r}_2(\tau) - \mathbf{r}_1(\tau)\right), \tag{4.76}$$

with ρ denoting the range between the two satellites. The vectors $\mathbf{r}_2(\tau)$ and $\mathbf{r}_1(\tau)$ describe the positions of the two GRACE satellites, and $\mathbf{e}_{12}$ is the unit vector in line-of-sight direction,

$$\mathbf{e}_{12}(\tau) = \frac{\mathbf{r}_{12}(\tau)}{\parallel \mathbf{r}_{12}(\tau) \parallel} \qquad \text{with} \qquad \mathbf{r}_{12}(\tau) = \mathbf{r}_2(\tau) - \mathbf{r}_1(\tau). \tag{4.77}$$

Analogous formulae can be derived for range-rate and range-acceleration measurements, as described in Mayer-Gürr (2006). Eq. (4.76) is a non-linear functional of the unknown parameters $\mathbf{x}$. In order to use it in a Gauss-Markoff model, it has to be linearized,

$$\rho = \rho_0 + \left.\frac{\partial \rho}{\partial \mathbf{x}}\right|_{\mathbf{x}_0} \cdot \Delta\mathbf{x} + ... \tag{4.78}$$

The relation between the range measurements and the unknown parameters is not stated explicitly in Eq. (4.76), but via the satellites' positions along the orbits. Thus the partial derivatives of the range measurements are derived by applying the chain rule and first differentiating with respect to the positions of the two satellites. Then the positions are differentiated with respect to the force function $\mathbf{f}$ and finally the force function with respect to unknown parameters,

$$\frac{\partial \rho}{\partial \mathbf{x}} = \frac{\partial \rho}{\partial \mathbf{r}_1} \cdot \frac{\partial \mathbf{r}_1}{\partial \mathbf{f}} \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \frac{\partial \rho}{\partial \mathbf{r}_2} \cdot \frac{\partial \mathbf{r}_2}{\partial \mathbf{f}} \cdot \frac{\partial \mathbf{f}}{\partial \mathbf{x}}. \tag{4.79}$$

The first term of each product is the differentiation of Eq. (4.76),

$$\frac{\partial \rho}{\partial \mathbf{r}_1} = -\mathbf{e}_{12}, \qquad \frac{\partial \rho}{\partial \mathbf{r}_2} = \mathbf{e}_{12}. \tag{4.80}$$

The second part of the partial derivatives, the linearized relationship between the positions and the unknown parameters $\partial\mathbf{r}/\partial\mathbf{x} = \partial\mathbf{r}/\partial\mathbf{f} \cdot \partial\mathbf{f}/\partial\mathbf{x}$, corresponds to the observation equations for the POD case, as given by Eq. (4.65). The observation equations for the range measurements are then obtained from the ones for the POD observations by multiplication with the derivatives in Eq. (4.80). This equals a projection onto the line of sight connection. It has to be pointed out that, in case of K-band observations, the refined satellite's positions given by Eq. (4.75) have to be introduced into the model, as the accuracy of the original GPS positions is not sufficient. The non-linear relation (4.76) requires a reference gravity field model to be introduced as approximation for the unknown parameters. Furthermore, the range observations have to be reduced by the influence of the other modelled forces, such as surface and tidal forces. The reduced observations are introduced as pseudo-observations on the left side of Eq. (4.76). For these pseudo-observations a covariance matrix can be determined, including the noise model of the satellite ranging system and of the accelerometer observations, see Mayer-Gürr (2006).

4.3.3 Satellite Gravity Gradiometry (GOCE)

The gradiometer instrument, as in case of the GOCE mission, determines the gravity gradient consisting of the second derivatives of the gravitational potential. The observation equation for these types of measurements is given by

$$\nabla\nabla V(\mathbf{r}) = \sum_{i=1}^{I} a_k \nabla\nabla\Phi_i(\mathbf{r}). \tag{4.81}$$

The design matrix $\mathbf{A}$ corresponds to the matrix $\mathbf{T}$ already described by Eq. (4.35),

$$\mathbf{A} = \mathbf{T}. \tag{4.82}$$

In this way, the matrix $\mathbf{A}$ contains the observation equations that have been rotated into the gradiometer reference frame. It should be pointed out that in practical observations not all of the nine tensor components will be measured with equal accuracy. On the contrary, the three components on the main diagonal of the gravity gradient tensor will be determined with superior accuracy. Therefore, those three components will serve as primary observations.

5. Solving the System of Observation Equations

The following chapter deals with the solution of the observation equations that are set up in Chapter 4.3. The solution is performed in the sense of a least squares adjustment, as introduced in Section 4.1. Particular interest is dedicated to the ill-posedness of the downward continuation process that is inevitable when dealing with data in satellite altitude and aiming at the determination of the gravity field on the Earth's surface., especially in case of space-localizing basis functions. Therefore, in the first part of this chapter, the concept and the implications of ill-posed problems are reviewed. To overcome the issues imposed by the downward continuation, the problem can be stabilized by the introduction of prior information. This corresponds to a regularization according to Tikhonov. Thus in the second part of this chapter, the regularization process is described for the applied case of a Tikhonov regularization. Subsequently, the implications of this kind of regularization are specified for the case of a parameterization by the space localizing basis functions used within this thesis. The regularization parameter is determined by a variance component estimation procedure, as described afterwards. Finally, the concept of a regionally adapted regularization will be introduced, as this plays a major role for the benefit of the regional gravity field recovery. In this sense, the regularization procedure presents an important module for the regional approach presented in this thesis.

5.1 Inverse Problems, Ill-posed Problems

When dealing with geodetic problems, especially in satellite geodesy, the concepts of inverse and ill-posed problems is of particular importance. The issue has been discussed widely within the geodetic community, see, for example, Moritz (1966), Schwarz (1971), Rummel et al. (1979), or Ilk (1984). A short introduction will be given in this section. An overview concerning this topic from the mathematical point of view can be found in Louis (1989) or Hansen (1997).

Let X and Y be normed vector spaces, and let $A : X \to Y$ be the physical model describing the linear mapping between the two spaces according to

$$\mathbf{A}\mathbf{x} = \mathbf{y}, \tag{5.1}$$

with $\mathbf{x} \in X$ and $\mathbf{y} \in Y$. The task of calculating the output $\mathbf{y}$ from a given set of $\mathbf{x}$ is described as direct problem, whereas the calculation of the model parameters indirectly from a given set of observations is specified as inverse problem. An inverse problem is called 'well posed' according to Hadamard (1923), as long as $\mathbf{A}$ is bijective, and the inverse operator $\mathbf{A}^{-1}$ is continuous. $\mathbf{A}$ being bijective means that the equation is solvable for all $\mathbf{y} \in Y$ (surjective), and that it has a unique solution (injective). Requiring the inverse operator to be continuous implicates that the solution $\mathbf{x}$ depends continuously on the data (with the concept of continuity introduced in Appendix A.2). This implies that small changes in $\mathbf{x}$ result in small changes in $\mathbf{y}$. In this case, the problem can be considered as being stable. In the context of a linear operator, the requirement of continuity is equivalent to assuming that the operator is bounded. As soon as any one of these requirements (existence, uniqueness, and stability) is violated, the problem is said to be an 'improperly posed' or 'ill-posed' problem. In geodesy one frequently has to deal with inverse problems, as the model parameters to be determined are usually not subject to direct observations. Inverse problems are frequently ill-posed, as very often at least one of the conditions mentioned above is not satisfied. An example of the non-uniqueness of the solution is the problem of determining the mass distribution in the interior of the Earth from the exterior gravitational potential, as infinitely many mass distributions can result in the same exterior gravity field. The reason for ill-posedness that will be further investigated in this thesis is the lack of stability. In case of satellite geodesy, the missing stability can have several causes, the most important of them being listed below:

- The calculation of gravity field functionals on the Earth's surface from measurements in satellite altitude causes an ill-posed problem, as during the downward continuation the signal is amplified especially in the high frequency part of the spectrum. This matter will be described in more detail in Section 5.1.1.
- Irregular data distribution or data gaps (such as the polar gap problem in case of the GOCE mission) lead to instabilities.
- A measurement instrument might not be able to recover the whole spectral domain. (For example, the GOCE gradiometer is not sensitive with respect to long wavelength features of the gravity field.)

5.1.1 Ill-posedness of the Downward Continuation Process

In this thesis, the focus in the context of ill-posed problems is on the downward continuation process in case of gravity field recovery from satellite data. Therefore, this problem shall be dealt with in more detail. The instability of this process can be observed by analyzing the compact upward continuation operator, as defined by Eq. (3.70),

$$U(\mathbf{x},\mathbf{x}_r) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \frac{1}{R} \left(\frac{R}{r}\right)^{n+1} Y_{nm}\left(\frac{\mathbf{x}_r}{|\mathbf{x}_r|}\right) Y_{nm}(\mathbf{x}). \tag{5.2}$$

This operator can be decomposed into its singular values, as described in Section A.2.0.1, with the orthogonal singular functions Y_{nm} and the singular values

$$\sigma_{nm} = \frac{1}{R}\left(\frac{R}{r}\right)^{n+1}. \tag{5.3}$$

Details concerning these spectral relationships can be found in MEISSL (1971) or RUMMEL and VAN GELDEREN (1995). The decay of the singular values towards zero with increasing frequency n becomes evident in Eq. (5.3). Thus high frequencies are associated with small singular values, which indicates the smoothing property of the kernel. This is a very important characteristic of compact operators, as defined by Eq. (A.37). A consequence of this characteristic is the fact that the inverse of a compact linear operator cannot be bounded. The inverse of the upward continuation operator is the downward continuation operator with its singular values approaching infinity with increasing n. This illustrates the amplification of small frequencies during the downward continuation process. As boundedness and continuity are equivalent concepts when dealing with linear operators, the downward continuation operator cannot be continuous. Thus it violates Hadamard's continuity assumption for well-posed problems.

5.1.2 Singular Value Decomposition of the Design Matrix

The instability of a problem given by Eq. (4.6) can be identified by an examination of the spectral behavior of the design matrix **A**. Therefore, the theory of spectral decomposition (of finite dimensional matrices) will shortly be reviewed in the following, as described, for example, by HANSEN (1997). For the case of infinite dimensional operators, the singular value decomposition is described in Appendix A.2.0.1. The concepts introduced there for infinite problems can help to understand the finite dimensional problem, as the design matrix **A** represents a discretization of the infinite dimensional operator.

The rectangular matrix $\mathbf{A} \in \mathbb{R}_{n\times u}$ with rank $n \geq u$ can be decomposed into

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{u} \mathbf{u}_i \sigma_i \mathbf{v}_i^T \qquad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_u \\ \mathbf{0} \end{pmatrix}, \tag{5.4}$$

with the orthogonal matrices $\mathbf{U} = (\mathbf{u}_1, ...\mathbf{u}_n) \in \mathbb{R}_{n\times n}$ and $\mathbf{V} = (\mathbf{v}_1, ...\mathbf{v}_u) \in \mathbb{R}_{u\times u}$. The matrix $\mathbf{\Sigma} \in \mathbb{R}_{n\times u}$ contains the diagonal matrix $\Sigma_u = \text{diag}(\sigma_1, ..\sigma_u)$ with the singular values σ_i that can be arranged in descending order according to

$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_u > 0. \tag{5.5}$$

The vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the corresponding left and right singular vectors, respectively, satisfying the following relations,

$$\begin{aligned} \mathbf{A}\mathbf{v}_i &= \sigma_i \mathbf{u}_i \\ \mathbf{A}^T \mathbf{u}_i &= \sigma_i \mathbf{v}_i. \end{aligned} \tag{5.6}$$

The singular values of the design matrix can be related to the eigenvalues of the normal equation matrix, as the $\mathbf{v}_i$ are the eigenvectors of $\mathbf{A}^T\mathbf{A}$ and the $\mathbf{u}_i$ the eigenvectors of $\mathbf{A}\mathbf{A}^T$ with the same eigenvalues λ_i for both matrices. The singular values are then given by $\sigma_i = \sqrt{\lambda_i}$. As described in Section 5.1.1, singular vectors corresponding to small singular values can be associated with high frequencies. Using this characteristic, the smoothing property of the matrix $\mathbf{A}$ can be observed when the mapping of an arbitrary vector $\mathbf{x}$ into the range $\mathcal{R}(\mathbf{A})$ (see Appendix A.2) is performed,

$$\mathbf{x} = \sum_{i=1}^{u} (\mathbf{x}^T \mathbf{v}_i)\mathbf{v}_i \qquad \mathbf{A}\mathbf{x} = \sum_{i=1}^{u} \sigma_i (\mathbf{x}^T \mathbf{v}_i)\mathbf{u}_i. \tag{5.7}$$

From the above Eq. (5.7), it becomes obvious that by the mapping these high frequencies of $\mathbf{x}$ are dampened more strongly than the lower frequencies due to the multiplication with smaller singular values σ_i. The opposite is the case as to the solution of the inverse problem,

$$\hat{\mathbf{x}} = \sum_{i=1}^{u} \frac{1}{\sigma_i} (\mathbf{y}^T \mathbf{u}_i)\mathbf{v}_i. \tag{5.8}$$

Eq. (5.8) can be considered as spectral decomposition of $\hat{\mathbf{x}}$ with the coefficients $\frac{1}{\sigma_i}(\mathbf{y}^T\mathbf{u}_i)$ indicating the spectral properties of $\hat{\mathbf{x}}$. The amplification of the smaller frequencies becomes evident. As formulated by Eq. (A.40), the observations must not be arbitrarily rough, as the coefficients $(\mathbf{y}^T\mathbf{u}_i)$ have to decay faster than the singular values σ_i, which imposes a smoothness condition on the observations. In case of a finite dimensional operator, (5.8) always converges. But the vector $\hat{\mathbf{x}}$ is the discretization of a continuous function which can only be continuous if, on the average, the frequency band decays with increasing frequency. Therefore, the requirement holds for discrete problems as well (Discrete Picard condition, for more details see, e.g., HANSEN 1997). Uncorrelated noise does not decline with higher frequencies, on the contrary, white noise is of equal magnitude for all frequencies, which poses a contradiction to the Picard condition.

The Picard condition is equivalent to demanding the observations $\mathbf{y}$ to be in the range $\mathcal{R}(\mathbf{A}) = \mathrm{span}(\mathbf{u}_1, ..., \mathbf{u}_u)$ of the matrix $\mathbf{A}$. But again, due to measurement noise present in the data, this cannot necessarily be assumed. The noise does not generally belong to $\mathcal{R}(\mathbf{A})$. As a solution to Eq (5.1) only exists for $\mathbf{y} \in \mathcal{R}(\mathbf{A})$, the system of equations cannot be solved in the rigorous way, but only the distance between $\mathbf{y}$ and $\mathbf{A}\mathbf{x}$ can be minimized. This implies that $\hat{\mathbf{x}}$ in Eq. (5.8) is the following solution in the least squares sense,

$$\hat{\mathbf{x}} = \mathbf{A}^+ \mathbf{y}, \tag{5.9}$$

with $\mathbf{A}^+$ denoting the generalized inverse or Moore-Penrose inverse, see, for example, GROETSCH (1977) or BJÖRCK (1996),

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \sum_{i=1}^{u} \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^T. \tag{5.10}$$

In case of compact operators in a Hilbert space, described in Appendix A.2, it is stated that a compact operator in the infinite dimensional case can never have a continuous and therefore bounded inverse. In case of a finite dimensional operator, if the inverse exists, it is also bounded and thus continuous. Therefore, strictly speaking, Hadamard's continuity principle is not violated. But, nevertheless, the solution becomes numerically unstable, and the amplified errors in the observations contaminate the solution.

A linear system of the form (5.1) is regarded as ill-posed if the singular values of $\mathbf{A}$ descend very fast. Crucial is the ratio of the largest singular value compared to the smallest singular value that is referred to as condition number of the matrix,

$$\mathrm{cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_u}. \tag{5.11}$$

In case of an ill-posed problem, it tends to take very large numbers. The matrix is then said to be ill conditioned. From Eq. (5.6) it can be assumed that singular values equal to zero span the null space $\mathcal{N}(\mathbf{A})$ of the matrix $\mathbf{A}$, as defined by Eq. (A.22). This results in the solution of Eq. (5.1) not to be unique any more. In case of singular values very close to zero, the system might be uniquely solvable from a strictly mathematical point of view, but will become severely unstable.

If the design matrix is ill conditioned, the data by itself is insufficient to obtain a stable solution. Referring to least squares estimation, this implies that the minimization of the square sum of the residuals,

$$J(\mathbf{x}) = \frac{1}{\sigma_\epsilon^2} (\mathbf{A}\mathbf{x} - \mathbf{y})^T \mathbf{P}_\epsilon (\mathbf{A}\mathbf{x} - \mathbf{y}) := \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2_{\mathcal{C}(\epsilon)}, \tag{5.12}$$

by itself does not lead to a stable solution. Therefore, additional prior information about the solution has to be introduced. These can be a-priori values for the unknown parameters, e.g. derived from existing models, and smoothness assumptions in terms of an a-priori given covariance matrix of the unknown parameters. The stabilization procedure is known as regularization and will be further described in the next section.

5.2 Regularization

In literature a variety of regularization methods has been proposed; an overview can be found, e.g., in Louis (1989), Bouman (1998), or Kusche (2002). Among them are the Tikhonov regularization (Tikhonov 1963), the truncated singular value decomposition (see, for example, Hansen 1987), and regularization techniques taking advantage of the regularizing character of iterative solution strategies (Landweber 1951) such as conjugate gradients. In this thesis, only the Tikhonov regularization has been applied, therefore it will be described in more detail in the following.

5.2.1 Tikhonov Regularization

This regularization method was independently developed by Tikhonov (1963) and Philips (1962) and is, therefore, also referred to as Tikhonov-Philips regularization. The procedure is based on the minimization of the functional (5.12) under consideration of a 'penalty term' $\|\mathbf{s}\|^2_K$ that describes the norm of the signal $\mathbf{s}$ given as linear functional $\mathbf{L}\mathbf{x}$ of the unknown parameters $\mathbf{x}$,

$$\mathbf{s} = \mathbf{L}\mathbf{x}. \tag{5.13}$$

$\|\mathbf{s}\|^2_K$ stands for a smoothing norm, for example given by the inner product of a reproducing kernel Hilbert space with kernel K according to (3.21),

$$\|\mathbf{s}\|^2_K = \langle \mathbf{s}, \mathbf{s} \rangle_K . \tag{5.14}$$

The requirement of this norm to be finite,

$$\|\mathbf{s}\|^2_K \leq c < \infty, \tag{5.15}$$

does not only introduce a bound on the norm of the solution, but does also imply a smoothness condition defined by the kernel K in equivalence with the smoothness assumption imposed by the kernel of an RKHS in Section 3.2. The resulting functional to minimize (5.12) and (5.14) simultaneously can be formulated as follows,

$$J_\alpha(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2_{\mathcal{C}(\epsilon)} + \alpha \|\mathbf{s}\|^2_K , \tag{5.16}$$

with α being the regularization or smoothing parameter. The minimization of Eq. (5.16) constitutes a compromise between minimizing the residual norm and keeping the penalty term $\|\mathbf{s}\|^2_K$ small. Due to the ill-posedness of the original model, small errors in the data lead to large deviations in the solution. In order to avoid this, the norm of the solution has to be limited.

5.2.2 Regularization in the Finite Dimensional Model

As smoothing kernel (in the following in the discrete sense defined by Eq. (3.21)) can serve an a-priori known covariance matrix of the signal,

$$\mathbf{K} = \mathcal{C}(\mathbf{s}) = \sigma_s^2 \mathbf{P}_s^{-1}, \tag{5.17}$$

with σ_s being the variance factor of the signal. The scalar product in an RKHS defined by a (finite dimensional) matrix can be formulated according to (3.21) as the $\mathcal{L}_2$ inner product containing the inverse of the kernel matrix following

$$\|\mathbf{s}\|^2_{\mathcal{C}(\mathbf{s})} = \langle \mathbf{s}, \mathbf{s} \rangle_{\mathcal{C}(\mathbf{s})} = \left\langle \mathbf{s}, \frac{1}{\sigma_s^2} \mathbf{P}_s \, \mathbf{s} \right\rangle. \tag{5.18}$$

Under consideration of the prior information

$$E(\mathbf{s}) = \mathbf{0} \qquad \text{and} \qquad \mathcal{C}(\mathbf{s}) = \sigma_s^2 \mathbf{P}_s^{-1}, \tag{5.19}$$

an equivalent formulation to (5.16) of the Tikhonov regularization based on the Bayes-approach is given by

$$J_\alpha(\mathbf{x}) = \left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{L} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \right\|^2_{\bar{\mathcal{C}}}, \tag{5.20}$$

with the inverse of the extended covariance matrix

$$\bar{\mathcal{C}}^{-1} = \begin{pmatrix} \frac{1}{\sigma_\epsilon^2} \mathbf{P}_\epsilon & 0 \\ 0 & \frac{1}{\sigma_s^2} \mathbf{P}_s \end{pmatrix} = \frac{1}{\sigma_\epsilon^2} \begin{pmatrix} \mathbf{P}_\epsilon & 0 \\ 0 & \frac{\sigma_\epsilon^2}{\sigma_s^2} \mathbf{P}_s \end{pmatrix}. \tag{5.21}$$

The expectation value $E(\mathbf{s}) = \mathbf{0}$ can be reasoned by the fact that (especially in a regional refinement approach) the estimated signal always represents a residual field to a (global) reference field. The minimum $\mathbf{x}_\alpha$ of (5.16) and (5.20) is defined by the unique solution of the regularized normal equations,

$$\begin{aligned} \mathbf{x}_\alpha &= \left[\begin{pmatrix} \mathbf{A}^T & \mathbf{L}^T \end{pmatrix} \begin{pmatrix} \mathbf{P}_\epsilon & 0 \\ 0 & \frac{\sigma_\epsilon^2}{\sigma_s^2} \mathbf{P}_s \end{pmatrix} \begin{pmatrix} \mathbf{A} \\ \mathbf{L} \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{A}^T & \mathbf{L}^T \end{pmatrix} \begin{pmatrix} \mathbf{P}_\epsilon & \mathbf{0} \\ \mathbf{0} & \frac{\sigma_\epsilon^2}{\sigma_s^2} \mathbf{P}_s \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} \\ &= (\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A} + \frac{\sigma_\epsilon^2}{\sigma_s^2} \mathbf{L}^T \mathbf{P}_s \mathbf{L})^{-1} \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y}. \end{aligned} \tag{5.22}$$

The regularization parameter α in Eq. (5.16) is interpreted here as signal-to-noise ratio,

$$\frac{1}{\alpha} = \frac{\sigma_s^2}{\sigma_\epsilon^2}. \tag{5.23}$$

This enables the regularized solution to be expressed according to

$$\mathbf{x}_\alpha = (\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A} + \alpha \mathbf{L}^T \mathbf{P}_s \mathbf{L})^{-1} \mathbf{A}' \mathbf{P}_\epsilon \mathbf{y}. \tag{5.24}$$

In this context, $\mathbf{x}_\alpha$ presents an unbiased least squares estimate of the unknown parameters under consideration of the prior information given in (5.19), when the Tikhonov regularization is interpreted as introduction of prior information in the Bayesian sense (KOCH 1990). It shall be mentioned, however, that the question whether the regularized solution can be regarded as unbiased is a controversially discussed matter, refer to, e.g., XU and RUMMEL (1994), XU et al. (2006), and KOCH and KUSCHE (1907).

The choice of the regularization parameter α is a crucial task, as it resembles the trade-off between the fitting of the solution to the given data set and the norm and smoothness of the regularized solution. The larger the regularization parameter is chosen, the stronger is the dampening and the smoother the solution. The parameter is to be chosen in accordance with the given signal. On the one hand, the solution has to be regularized as strongly as necessary to obtain a reasonable solution. On the other hand, the signal should not be dampened too much in order not to lose any information contained in the data. Several procedures to choose this parameter have been proposed, e.g. the L-curve criterion (HANSEN 1992), the generalized cross validation (going back to WAHBA 1977, see also HANSEN 1987), minimizing the total mean square error (XU 1992), or the discrepancy principle (MOROZOV 1966). In this thesis, the regularization paramater is interpreted as signal-to-noise ratio and is determined by variance component estimation as proposed by KOCH and KUSCHE (2001). Details of this method are given in Section 5.2.4.

5.2.3 Regularization and Splines

As described in Chapter 5.2, the regularization procedure aims at smoothing the solution of an ill-posed problem. In order to do so, certain smoothing conditions are imposed on the solution by adding a penalty term to the minimization functional according to (5.16). This penalty term $\|\mathbf{s}\|_K^2$ restricts the norm of a linear functional of the unknown parameters, in this case the spline coefficients a_i being arranged in the vector $\mathbf{x} = (a_0, ..., a_I)^T$. The linear functional $\mathbf{Lx}$ of the unknown parameters can then be interpreted as

$$\mathbf{Lx} = \mathbf{s}(\mathbf{x}) = \sum_{i=1}^{I} a_i \Phi_i(\mathbf{x}, \mathbf{x}_i). \tag{5.25}$$

Considering the a-priori information given in (5.19) and taking (5.17) into account, the functional (5.16) can be reformulated according to

$$J_\alpha(\mathbf{x}) = (\mathbf{Ax} - \mathbf{y})^T \mathbf{P}_\epsilon (\mathbf{Ax} - \mathbf{y}) + \alpha \langle \mathbf{s}, \mathbf{s} \rangle_{\mathcal{C}}. \tag{5.26}$$

The smoothing condition is induced by using the a-priori covariance matrix of the signal $\mathcal{C} = \mathcal{C}(\mathbf{s})$, in case of the gravitational potential given by (3.58). With the signal expressed in terms of radial basis functions according to Eq.(3.74), the scalar product is defined by

$$\langle \mathbf{s}, \mathbf{s} \rangle_{\mathcal{C}} = \left\langle \sum_i a_i \Phi_i, \sum_k a_k \Phi_k \right\rangle_{\mathcal{C}} = \sum_i \sum_k a_i a_k \langle \Phi_i, \Phi_k \rangle_{\mathcal{C}}. \tag{5.27}$$

If the scalar products of the basis functions $\langle \Phi_i, \Phi_k \rangle_{\mathcal{C}}$ are combined in the regularization matrix $\mathbf{R}$ with the elements

$$R_{ik} = \langle \Phi_i, \Phi_k \rangle_{\mathcal{C}}, \tag{5.28}$$

then (5.26) becomes

$$J_\alpha(\mathbf{x}) = (\mathbf{Ax} - \mathbf{y})^T \mathbf{P}_\epsilon (\mathbf{Ax} - \mathbf{y}) + \alpha \cdot \mathbf{x}^T \mathbf{Rx}. \tag{5.29}$$

The regularized solution for the minimization of the functional $J_\alpha(\mathbf{x})$ is then given by

$$\mathbf{x}_\alpha(\mathbf{x}) = (\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A} + \alpha \mathbf{R})^{-1} \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y}. \tag{5.30}$$

Next, the nature of the regularization matrix $\mathbf{R}$ shall be investigated. Therefore, the elements of the regularization matrix in (5.28) will be calculated. In the limit, the inner products represent integrals over the sphere. If the covariance function (3.58) is chosen as reproducing kernel, and the basis functions are chosen according to (3.73), the elements of the regularization matrix result in

$$\begin{aligned} R_{i,k} = \langle \Phi_i, \Phi_k \rangle_{\mathcal{C}} &= \iint_{\Omega} \frac{1}{k_n^2} \left[\sum_{n=2}^{\infty} \sum_{m=-n}^{n} k_n Y_{nm}(\mathbf{x}) Y_{nm}(\mathbf{x}_i) \right] \left[\sum_{\bar{n}=2}^{\infty} \sum_{\bar{m}=-\bar{n}}^{\bar{n}} k_{\bar{n}} Y_{\bar{n}\bar{m}}(\mathbf{x}) Y_{\bar{n}\bar{m}}(\mathbf{x}_k) \right] \\ &= \sum_{n=2}^{\infty} \sum_{m=-n}^{n} Y_{nm}(\mathbf{x}_i) Y_{nm}(\mathbf{x}_k) = \delta(\mathbf{x}_i, \mathbf{x}_k). \end{aligned} \tag{5.31}$$

$\delta(\mathbf{x}_i, \mathbf{x}_k)$ denotes the Dirac functional, which implies that it has the value zero for two different functions, but is infinite in the case $i = k$ with the integral over δ being one. Thus in a certain sense, the basis functions diagonalize the regularization matrix, as any two functions located at two different points are decorrelated with respect to the inner product defined by the kernel. This is very much appreciated, as it allows the separation of the regularization matrix, as will be described in Section 5.2.5. But the problem is the fact that each Φ_i itself does not possess finite energy concerning the scalar product defined by $\mathcal{C}$. This means that the norm with respect to $\mathcal{C}$ is not finite,

$$\|\Phi_i\|_{\mathcal{C}} = \infty, \tag{5.32}$$

which results in the elements on the main diagonal of $\mathbf{R}$ to become infinite. The implication of Eq. (5.32) is the fact that the basis functions Φ_i do not belong to the RKHS defined by the covariance function $\mathcal{C}$. They are not smooth enough to satisfy the requirement imposed by $\mathcal{C}$ as reproducing kernel, as has already been stated in Eq. (3.90)

If one is interested in adapting the basis functions in such a way to make them part of the Hilbert space defined by the reproducing kernel $\mathcal{C}$, the basis functions would have to become smoother. This could be achieved by introducing a damping factor of $(1/\sqrt{1+\epsilon})^n$ (ϵ being a small positive number) into the series expansion,

$$\bar{\Phi}_i(\mathbf{x}, \mathbf{x}_i) = \sum_{n=2}^{\infty} \sqrt{2n+1} \cdot k_n \cdot \left(\frac{1}{\sqrt{1+\epsilon}}\right)^n P(\mathbf{x} \cdot \mathbf{x}_i), \tag{5.33}$$

leading to the following elements of the regularization matrix,

$$R_{i,k} = \langle \bar{\Phi}_i, \bar{\Phi}_k \rangle_{\mathcal{C}} = \sum_{n=2}^{\infty} \sum_{m=-n}^{n} \left(\frac{1}{1+\epsilon}\right)^n Y_{nm}(\mathbf{x}_i) Y_{nm}(\mathbf{x}_k). \tag{5.34}$$

Eq. (5.34) reveals that, in contrast to the original basis functions, the modified spline kernels lack the orthogonality regarding the respective inner product.

The above considerations have been made for non-bandlimited spline functions. In practical calculations, however, the respective basis functions always have to be bandlimited with an upper degree N leading to the functions defined in Section 3.4.3. From this it follows that the series expansion used in the calculation of the scalar product in (5.31) is truncated at degree N as well. When dealing with a truncated series, of course, the problem of infinite energy does not exist any more. Hence, in case of bandlimited spline functions, the energy on the main diagonal of the regularization matrix becomes finite, even without introduction of a damping factor. However, the inner product (5.31) can only be orthogonal if the summation is performed up to $N = \infty$. Thus the bandlimited spline functions lose their strict orthogonality in the sense of the inner product with respect to $\mathcal{C}$. The elements R_{ik} of the regularization matrix reflect the correlations between two different basis functions located at two different nodal points $\mathbf{x}_i$ and $\mathbf{x}_k$.

An ideal case would provide a set of basis functions being decorrelated by the scalar product induced by the covariance function as reproducing kernel, but at the same time having finite energy with respect to the norm defined by the same kernel. The simultaneous satisfaction of both requirements would result in the regularization matrix becoming the unit matrix (or at least differing from the unit matrix only by a constant factor which could be absorbed in the regularization parameter). However, the fulfillment of both conditions is not easily available (especially for bandlimited basis functions) due to the issues of convergence of the scalar product and the loss of orthogonality discussed above.

Nevertheless, the regularization matrix is to be approximated by the unit matrix. This provides the advantage of separating the regularization matrix and calculating different matrices for different areas, as will be discussed in more detail in Section 5.2.5. The approximation

$$\mathbf{R} = \mathbf{I} \tag{5.35}$$

leads to the functional $J_\alpha(\mathbf{x})$ of Eq. (5.29) to be modified according to

$$J_\alpha(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{y})^T \mathbf{P}_\epsilon (\mathbf{A}\mathbf{x} - \mathbf{y}) + \alpha \cdot \mathbf{x}^T \mathbf{x}, \tag{5.36}$$

resulting in the regularized solution

$$\mathbf{x}_\alpha(\mathbf{x}) = (\mathbf{A}^T \mathbf{P}_\epsilon \mathbf{A} + \alpha \mathbf{I})^{-1} \mathbf{A}^T \mathbf{P}_\epsilon \mathbf{y}. \tag{5.37}$$

It is understood that the use of $\mathbf{I}$ instead of $\mathbf{R}$ is only an approximation, therefore it has to be investigated in how far this approximation has any effect on the solution. This will be attended to in the following considerations and in a simulation scenario described in Section 7.2.

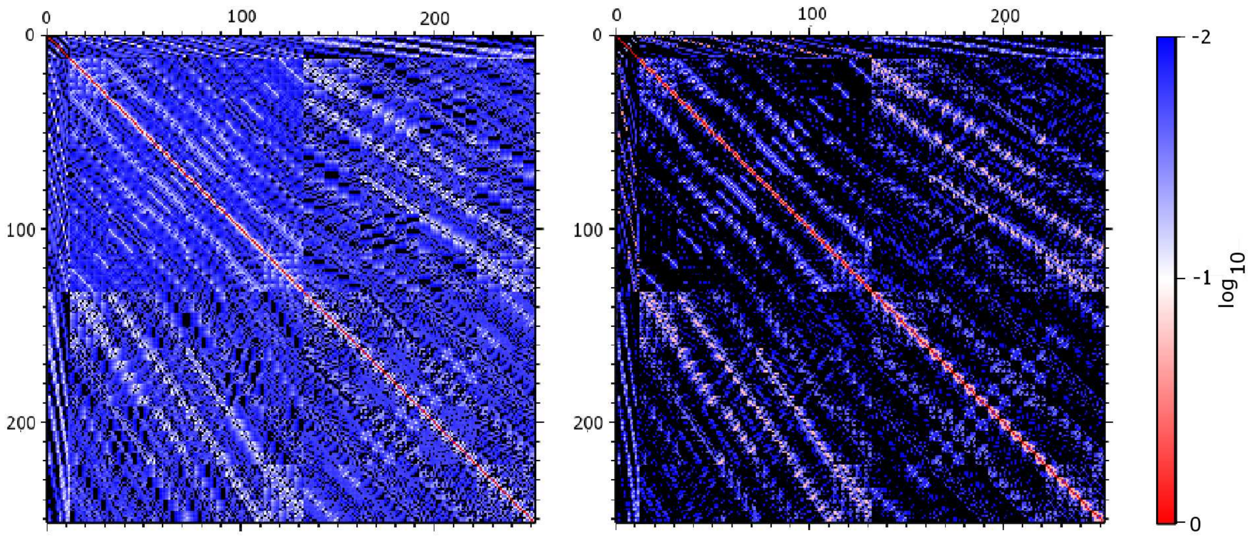


Figure 5.1: Regularization matrix calculated as the inner products of basis functions, left: original spline kernel, right: modified spline kernel (modification factor 0.9), basis functions developed up to $N = 14$, located on a triangle vertex grid of level 4

To illustrate the effect of the non-orthogonality of the bandlimited spline kernels and to show the effect of the modified basis functions (5.33), examples of regularization matrices are plotted in Fig. 5.1. The spline kernels in this example have been developed up to $N = 14$ and have been arranged on a matching triangle vertex grid (see Section 3.5.1.6) of level 4. The left hand side of the figure shows the regularization matrix resulting from the (bandlimited) inner products for the unmodified basis functions given by (5.31). The matrix is normalized to provide the value one for the elements on the main diagonal. The figure specifies that the elements apart from the main diagonal are significantly smaller than one, the highest correlation between two different spline kernels being about 0.12, as listed in Tab. 5.1. Even though this represents only one example of a specific basis function and grid combination, the situation does not change much when the regularization matrix is calculated for higher resolution spline kernels and their corresponding nodal point arrangements. The highest correlation factor is always around 10% of the value on the main diagonal. On the right hand side of Fig. 5.1, the same regularization matrix is plotted for modified spline kernels and their inner products given by (5.34), here presented with a modification factor of $(1/\sqrt{1+\epsilon})^n = 0.9$. This again represents only an exemplary value, the more $\epsilon \to 0$, the more the regularization matrix resembles the original one. While the overall sum of the absolute values of the elements apart from the main diagonal decreases with the modification factor, the highest correlation and the standard deviation of the off-diagonal elements increase. The impact of the modification factor on the basis functions and the resulting regularization matrices will be investigated in a simulation scenario in Section 7.2. There the impact of the unit matrix to be used as regularization matrix will be investigated, too.

	highest correlation	standard deviation	sum of absolute values
original kernel	0.128337	6.08808	1163.07
modified kernel (factor 0.9)	0.285343	9.01617	921.36

Table 5.1: Difference between the regularization matrix and the unit matrix, triangle vertex level 4, $N = 14$

Mathematically, the impact of the use of the unit matrix as regularization matrix can be exploited by investigating the difference $\mathbf{D}$ between regularized normal equations with fully occupied regularization matrix compared to the normal equations with the use of the unit matrix according to

$$\mathbf{D} = (\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{I})^{-1} - (\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{R})^{-1}. \tag{5.38}$$

With the simplifications $\mathbf{B} = (\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{I})$ and $\mathbf{B}' = (\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{R})$, this reads

$$\mathbf{D} = \mathbf{B}^{-1} - \mathbf{B}'^{-1}. \tag{5.39}$$

Factoring out the matrices $\mathbf{B}'^{-1}$ and $\mathbf{B}^{-1}$ results in

$$\mathbf{D} = \mathbf{B}^{-1}(\mathbf{I} - \mathbf{B}\mathbf{B}'^{-1}) = \mathbf{B}^{-1}(\mathbf{B}' - \mathbf{B})\mathbf{B}'^{-1}. \tag{5.40}$$

Resubstituting the original matrix expressions results in the following difference,

$$\mathbf{D} = (\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{I})^{-1}(\mathbf{R} - \mathbf{I})(\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{R})^{-1}. \tag{5.41}$$

This implies that the difference between the regularized normal equations can directly be expressed by the difference $(\mathbf{R} - \mathbf{I})$ between the matrices themselves.

It shall be pointed out that a regularization matrix equivalent to Eq. (5.34) would be obtained in case of unmodified basis functions if the modified covariance function $\bar{\mathcal{C}}$, proposed in Eq. (3.69), was applied

$$\left\langle \bar{\Phi}_i, \bar{\Phi}_k \right\rangle_{\mathcal{C}} = \left\langle \Phi_i, \Phi_k \right\rangle_{\bar{\mathcal{C}}}. \tag{5.42}$$

In this case, the basis functions (even without damping factor) would directly belong to the RKHS defined by $\bar{\mathcal{C}}$. Even though the regularization matrix equals the one given in Eq. (5.34), the basis functions used in the modeling differ in both cases. The impact of this modified covariance matrix and the resulting regularization matrix is also investigated in the simulation scenario presented in Section 7.2.

5.2.4 Variance Component Estimation

If different kinds of observations are to be combined, the determination of the correct relative weighting of the observations is essential to receive a correct result. In the same context, the choice of the regularization parameter can be treated if the regularization is interpreted as prior information in the Bayesian sense. For both tasks the method of variance component estimation (VCE), as described by Koch and Kusche (2001), can be applied.

The solution can be estimated from a system of (combined) normal equations $\mathbf{N}$ which is accumulated as a weighted sum of the normal equation systems $\mathbf{N}_k$ of the individual observation groups. This can be formulated according to

$$\mathbf{N}\hat{\mathbf{x}} = \mathbf{n} \qquad \text{with} \quad \mathbf{N} = \sum_k \frac{1}{\sigma_k^2}\mathbf{N}_k \quad \text{and} \quad \mathbf{n} = \sum_k \frac{1}{\sigma_k^2}\mathbf{n}_k. \tag{5.43}$$

The weighting factors are the reciprocal variances of the normal equations,

$$\hat{\sigma}_k^2 = \frac{\Omega_k}{r_k}, \tag{5.44}$$

with

$$\Omega_k = \hat{\mathbf{e}_k}^T\mathbf{P}_k\hat{\mathbf{e}_k} = (\mathbf{A}_k\hat{\mathbf{x}} - \mathbf{l}_k)^T\mathbf{P}_k(\mathbf{A}_k\hat{\mathbf{x}} - \mathbf{l}_k), \tag{5.45}$$

being the square sum of the residuals of the kth group of observations and

$$r_k = n_k - \frac{1}{\sigma_k^2}\mathrm{tr}\left(\mathbf{N}_k\mathbf{N}^{-1}\right) \tag{5.46}$$

their partial redundancy with n_k denoting the number of observations in the kth group. The partial redundancies sum up to the overall redundancy $\sum_k r_k = n - u$. The solution $\hat{\mathbf{x}}$ and the variances σ_k^2 are unknown a-priori, consequently an iterative procedure is inevitable, as illustrated in the flow chart of Fig. 5.2.

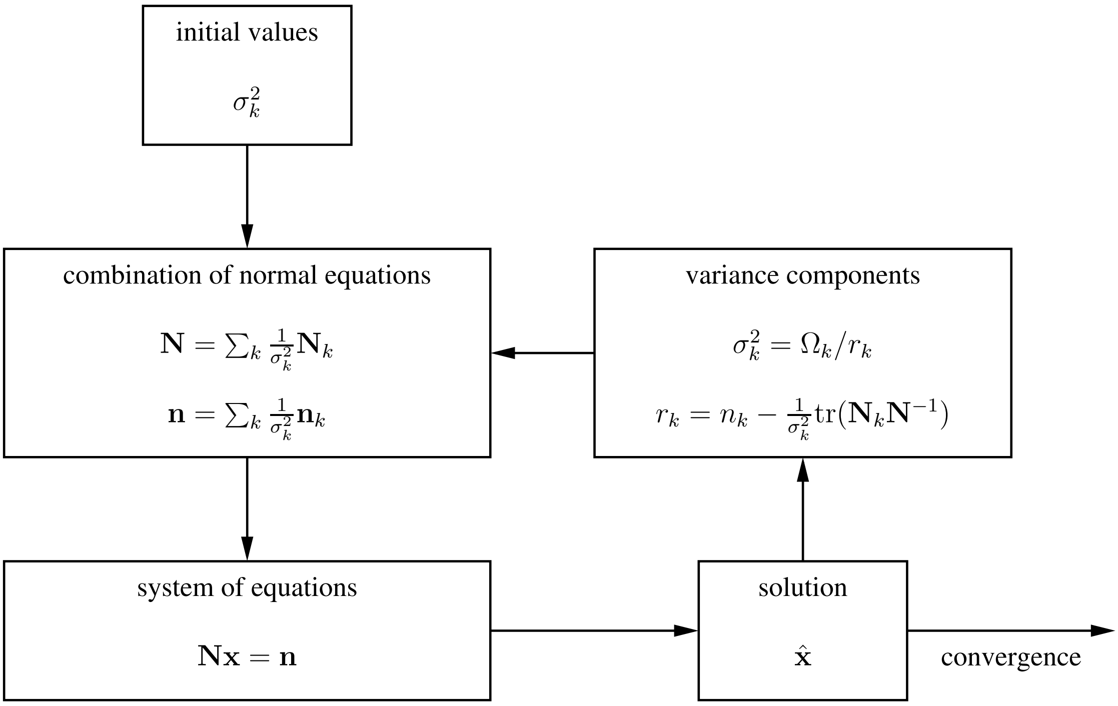


Figure 5.2: Iterative procedure of variance component estimation

As in this thesis the method of VCE is primarily utilized for the task of the determination of the regularization parameter, this special case of the Tikhonov regularization ought to be stated explicitly as well. Here the described approximation of the regularization matrix by the unit matrix is assumed, which leads to the following system,

$$\mathbf{N} = \frac{1}{\sigma_\epsilon^2}\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \frac{1}{\sigma_s^2}\mathbf{I}, \quad \text{and} \quad \mathbf{n} = \frac{1}{\sigma_\epsilon^2}\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{l}, \tag{5.47}$$

with σ_ϵ being the standard deviation of the observations and σ_s being the standard deviation of the signal. Both variance components can be treated as above described.

5.2.5 Regionally Adapted Regularization

The process of variance component estimation delivers the optimal regularization parameter under consideration of the given signal-to-noise ratio. In case of a regional gravity field determination, this results in one regularization parameter tailored optimally to the respective recovery region. This is an improvement in comparison to a global gravity field parameterization which allows only one regularization factor for the complete Earth, resulting in an overall mean damping of the gravity field features. But even within smaller geographical areas, the gravity field features may vary significantly. Therefore, it seems reasonable to further adapt the regularization procedure. The proposed approach does not take into account only one regularization matrix with one associated regularization parameter per region, but allows several matrices with respective parameters,

$$\mathbf{N} = \frac{1}{\sigma_\epsilon^2}\mathbf{A}^T\mathbf{P}\mathbf{A} + \frac{1}{\sigma_{s_1}^2}\mathbf{R}_1 + ... + \frac{1}{\sigma_{s_n}^2}\mathbf{R}_n. \tag{5.48}$$

The original regularization matrix is, therefore, split up into individual regularization matrices $\mathbf{R}_i$, each belonging to a regional regularization area i. To separate the original regularization matrix, the approximation by a unit matrix (5.35) made above is very convenient. $\mathbf{R}$ can only be separated that easily if the basis

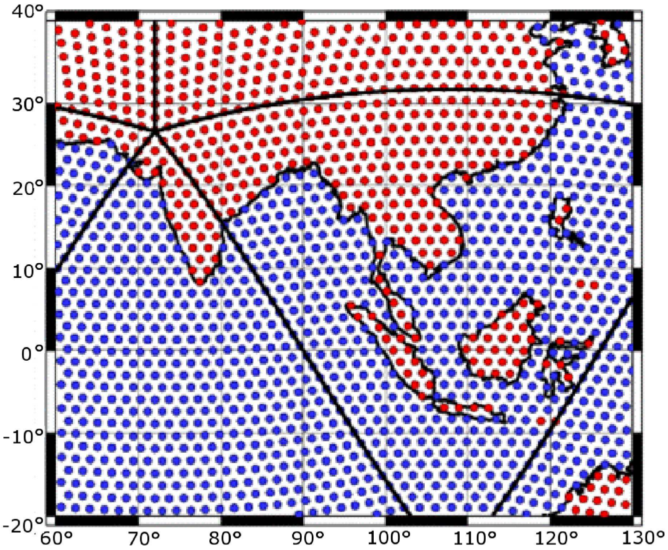


Figure 5.3: Two different regularization areas and the locations of the unknown parameters belonging to each area

functions are assumed to be orthogonal with respect to the inner product given by (5.31). In this case, each individual regularization matrix $\mathbf{R}_i$ is a diagonal matrix that features a 'one' for each regional spline parameter located inside the corresponding region and a 'zero' for parameters belonging to basis functions outside the regularization group,

$$\mathbf{R}_i(j,j) = \begin{cases} 1 & \text{for} \quad j \text{ inside } i \\ 0 & \text{for} \quad j \text{ outside } i \end{cases} . \tag{5.49}$$

The original identity matrix, as applied in the Tikhonov regularization process, has thus been divided into single diagonal matrices according to the membership of the respective unknown parameters to the different regularization groups,

$$\mathbf{R}_1 + ... + \mathbf{R}_n = \mathbf{I}. \tag{5.50}$$

The possibility of adapting the regularization procedure in this particular way is a unique feature of a field parameterization by space localizing basis functions, as each unknown parameter is related to a particular geographical location. This is an inevitable premise when the elements of the regularization matrix are supposed to be assigned to a certain region. The separation of a geographical region into different regularization areas is exemplarily illustrated in Fig. 5.3. Here the two regularization areas are the continental and the oceanic regions, resulting in the following normal equation matrix,

$$\mathbf{N} = \frac{1}{\sigma_\epsilon^2}\mathbf{A}^T\mathbf{P}\mathbf{A} + \frac{1}{\sigma_{Land}^2}\mathbf{R}_{Land} + \frac{1}{\sigma_{Ocean}^2}\mathbf{R}_{Ocean}. \tag{5.51}$$

The separation into land and ocean areas can propose a reasonable choice in certain regions, where the gravity field information on the oceans is significantly less rough compared to the continent areas. In the presence of deep sea trenches or rough ocean bottom topography, however, this might not always be a valid assumption. In this case, different choices for the regularization areas are inevitable.

5.3 Relationship Between Spline Approximation and Collocation

In the following, the relationships shall be investigated that exist between the (regularized) least squares solution parameterized by the space localizing basis functions defined in Section 3.4.2 and the least squares collocation approach. The considerations will be performed without going into detail about the concept of least squares collocation; information about this method can be found in the literature, examples are MORITZ (1962), KRARUP (1969), and MORITZ (1978).

The following formula specifies the case of least squares collocation with noise. It describes the prediction of a signal $\mathbf{s}$ on the basis of a set of observations $\mathbf{l}$,

$$\mathbf{s} = \mathbf{C}_{s,l}\left(\mathbf{C}_{l,l} + \alpha \mathbf{P}_\epsilon^{-1}\right)^{-1}\mathbf{l}, \tag{5.52}$$

with the autocovariance matrix $\mathbf{C}_{l,l}$ of the signal part of the observations, the cross-covariance matrix $\mathbf{C}_{s,l}$ between the signal, and the observations and the covariance matrix $\mathbf{P}$ of the observation noise. The factor α can be regarded as signal-to-noise ratio,

$$\frac{1}{\alpha} = \frac{\sigma_x}{\sigma_\epsilon}. \tag{5.53}$$

The solution obtained by least squares collocation is optimal in the sense of minimal variance on the basis of the given observations (MORITZ 1980). In the following, it will be assumed that the observations as well as the functionals to be predicted at different positions are given as gravitational potential. The same prediction as given in Eq. (5.52) can be performed by using the spline representation,

$$\mathbf{s} = \mathbf{B}\mathbf{x}. \tag{5.54}$$

The vector $\mathbf{x}$ contains the estimated spline parameters a_i given in Eq. (3.74), and $\mathbf{B}$ represents the prediction matrix according to Eq. (3.76) with the elements

$$(B)_{ki} = \sum_{n=2}^{N} \sigma_n \cdot P_n(\cos(\mathbf{x}_k \cdot \mathbf{y}_i)) = \sum_{n=2}^{N}\sum_{m=-n}^{n} \frac{\sigma_n}{\sqrt{2n+1}} Y_{nm}(\mathbf{y}_k) Y_{nm}(\mathbf{x}_i). \tag{5.55}$$

Here the $\mathbf{y}_k$ are the positions to which the signal is predicted, while $\mathbf{x}_i$ denote the nodal points of the spline kernels. By substituting the solution of the least squares adjustment for the estimated parameters in Eq.(5.54), the following relationship is obtained,

$$\mathbf{s} = \mathbf{B}\left(\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{A} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{P}_\epsilon\mathbf{l}. \tag{5.56}$$

Rearranging Eq, (5.56) by making use of matrix identities as given in, for example, KOCH (1997) yields

$$\mathbf{s} = \mathbf{B}\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T + \alpha\mathbf{P}_\epsilon^{-1}\right)^{-1}\mathbf{l}. \tag{5.57}$$

Comparing Eq. (5.52) and Eq. (5.57) reveals that the two formulas exhibit certain similarities. They would be identical if the following relationships would held,

$$\mathbf{B}\mathbf{A}^T = \mathbf{C}_{s,l} \qquad \text{and} \qquad \mathbf{A}\mathbf{A}^T = \mathbf{C}_{l,l}. \tag{5.58}$$

Carrying out the matrix multiplications results for the individual matrix elements in

$$\begin{aligned}
\mathbf{B}\mathbf{A}^T_{(kj)} &= \sum_{i=1}^{I}\left(\sum_{n=2}^{N}\sum_{m=-n}^{n}\frac{\sigma_n}{\sqrt{2n+1}}Y_{nm}(\mathbf{y}_k)Y_{nm}(\mathbf{x}_i)\right)\left(\sum_{\bar{n}=2}^{N}\sum_{\bar{m}=-\bar{n}}^{\bar{n}}\frac{\sigma_{\bar{n}}}{\sqrt{2\bar{n}+1}}Y_{\bar{n}\bar{m}}(\mathbf{y}_j)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)\right) \\
&= \sum_{n=2}^{N}\sum_{m=-n}^{n}\sum_{\bar{n}=2}^{N}\sum_{\bar{m}=-\bar{n}}^{\bar{n}}\frac{\sigma_n\sigma_{\bar{n}}}{\sqrt{2n+1}\sqrt{2\bar{n}+1}}Y_{nm}(\mathbf{y}_k)Y_{\bar{n}\bar{m}}(\mathbf{y}_j) \quad \underbrace{\sum_{i=1}^{I}Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)}_{\approx\int_\Omega Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)d\Omega=\delta_{n\bar{n}}\delta_{m\bar{m}}} \\
&\approx \sum_{n=2}^{N}\sum_{m=-n}^{n}\frac{\sigma_n^2}{2n+1}Y_{nm}(\mathbf{y}_k)Y_{nm}(\mathbf{y}_j) = \mathbf{C}_{s,l(kj)},
\end{aligned} \tag{5.59}$$

and

$$\begin{aligned}
\mathbf{AA}^T_{(hj)} &= \sum_{i=1}^{I}\left(\sum_{n=2}^{N}\sum_{m=-n}^{n}\frac{\sigma_n}{\sqrt{2n+1}}Y_{nm}(\mathbf{y}_h)Y_{nm}(\mathbf{x}_i)\right)\left(\sum_{\bar{n}=2}^{N}\sum_{\bar{m}=-\bar{n}}^{\bar{n}}\frac{\sigma_{\bar{n}}}{\sqrt{2\bar{n}+1}}Y_{\bar{n}\bar{m}}(\mathbf{y}_j)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)\right)\\
&= \sum_{n=2}^{N}\sum_{m=-n}^{n}\sum_{\bar{n}=2}^{N}\sum_{\bar{m}=-\bar{n}}^{\bar{n}}\frac{\sigma_n\sigma_{\bar{n}}}{\sqrt{2n+1}\sqrt{2\bar{n}+1}}Y_{nm}(\mathbf{y}_h)Y_{\bar{n}\bar{m}}(\mathbf{y}_j)\quad\underbrace{\sum_{i=1}^{I}Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)}_{\approx\int\limits_{\Omega}Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)d\Omega=\delta_{n\bar{n}}\delta_{m\bar{m}}} \qquad (5.60)\\
&\approx \sum_{n=2}^{N}\sum_{m=-n}^{n}\frac{\sigma_n^2}{2n+1}Y_{nm}(\mathbf{y}_h)Y_{nm}(\mathbf{y}_j)=\mathbf{C}_{l,l(hj)},
\end{aligned}$$

with $\mathbf{y}_h$ and $\mathbf{y}_j$ denoting the observation points. A comparison with Eq. (3.58) allows the assumption that the matrix products $\mathbf{BA}^T$ and $\mathbf{AA}^T$ in Eq. (5.57) show certain resemblances to the covariance matrices $\mathbf{C}_{s,l}$ and $\mathbf{C}_{l,l}$. This is a consequence of the shape coefficients of the radial basis functions k_n in Eq. (3.73) being chosen according to Eq. (3.77) as the square root of the coefficients of the covariance function. This can be regarded as evidence that the choice of the shape coefficients seems to be a reasonable one. Nevertheless, Eqs. (5.59) and (5.60) only represent approximations. The matrix products would only become exactly the covariance matrices if the sum over the basis functions $\sum_{i=1}^{I}Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)$ would turn into the integral over the sphere $\int\limits_{\Omega}Y_{nm}(\mathbf{x}_i)Y_{\bar{n}\bar{m}}(\mathbf{x}_i)d\Omega$. This would require an infinitely dense distribution of spline kernels over the surface of the sphere. Since this is not the case in practical calculations, no one-to-one relationship between the spline approximation as treated in this thesis and the least squares collocation can be derived from the above considerations. However, they confirm the usefulness of the basis functions. Further investigations of the discussed relationships are necessary.

6. From Regional to Global Gravity Fields

For a wide variety of applications, the calculation of regional gravity field solutions meets exactly the given requirements. Nevertheless, for some other applications, it seems to be useful to derive a global gravity field model by spherical harmonics without losing the details of a regional zoom-in. In the geodetic context, a set of spherical harmonics is often required, as it can easily be distributed and generally be handled without further instructions. Furthermore, it might be reasonable to deliver a gravity field parameterized by spherical harmonics for purposes of comparison. Most of the existing gravity field models are provided in terms of spherical harmonic coefficients, and, for instance, comparisons of accuracy are frequently performed on the basis of error degree variances. Thus if the gravity field model is to be published and acknowledged by the geodetic user community, the allocation of a spherical harmonic expansion appears to be desirable. In the first section of this chapter, it is derived how the spherical harmonic coefficients can directly be calculated from the coefficients of a global model parameterized by radial basis functions. If individual regional gravity field solutions are available with global coverage, then the determination of the spherical harmonic coefficients can be performed by means of quadrature methods, as described in the second section of this chapter.

6.1 Conversion from a Global Spline Representation to Spherical Harmonics

In principle, a gravity field representation based on space localizing basis functions cannot only be used for regional models, but for the representation of global gravity fields as well. The spline kernels are then located at a global grid, and a global set of observations is needed. Such a global spline representation can be transformed into a series of spherical harmonics without loss of information. This can be understood from the following considerations. Starting point is a representation of the gravitational potential in terms of splines,

$$
\begin{aligned}
V(\mathbf{r}) = \sum_{i=1}^{I} a_i \Phi_i(\mathbf{r}) &= \sum_{i=1}^{I} a_i \left[\frac{GM}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sqrt{2n+1}\; k_n \, P_n\left(\frac{\mathbf{r}}{r} \cdot \frac{\mathbf{r}_i}{R}\right) \right] \\
&= \sum_{i=1}^{I} a_i \left[\frac{GM}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} k_n \sum_{m=-n}^{n} Y_{nm}\left(\frac{\mathbf{r}}{r}\right) Y_{nm}\left(\frac{\mathbf{r}_i}{R}\right) \right] . \qquad (6.1)
\end{aligned}
$$

Eq. (6.1) can be reordered to yield

$$
V(\mathbf{r}) = \frac{GM}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} \left(\sum_{i=1}^{I} a_i k_n \, Y_{nm}\left(\frac{\mathbf{r}_i}{R}\right) \right) Y_{nm}\left(\frac{\mathbf{r}}{r}\right) . \qquad (6.2)
$$

Comparing this with the expansion of the gravitational potential in terms of spherical harmonics according to Eq. (3.42),

$$
V(\mathbf{r}) = \frac{GM}{R} \sum_{n=2}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=-n}^{n} c_{nm} Y_{nm}\left(\frac{\mathbf{r}}{r}\right) , \qquad (6.3)
$$

the spherical harmonic coefficients can be computed directly from the spline coefficients according to

$$
c_{nm} = \sum_{i=1}^{I} a_i \, k_n \, Y_{nm}\left(\frac{\mathbf{r}_i}{R}\right) . \qquad (6.4)
$$

6.2 Patching of Individual Regional Solutions

If regional solutions have been calculated independently, the determination of a spherical harmonic expansion cannot simply be performed by applying Eq. (6.4). In order to avoid truncation effects, the satellite data has to be taken in an area slightly larger than the evaluation area itself. The radial basis functions are not perfectly space localizing, therefore those located in the boundary area have to be taken into account as well, as they have influence on the gravity field values in the inner evaluation area. The resulting global solution, however, is only to be composed of the inner areas without the boundaries. Therefore, the patching of the regional solutions has to be based on perfectly localizing values, and the strategy described by Eq. (6.4) does not work for individually calculated solutions. Thus a different strategy has to be used. The concept that has been applied in this thesis is the merging of the regional solutions to a global one by means of quadrature methods, as introduced in EICKER et al. (2004). From the regional solutions, the gravity functionals, parameterized by the spline representation in the specific regions, can be calculated at the nodes of a global grid. This can be performed, in principle with arbitrary resolution, and the continuous modeling by spherical splines allows the functionals to be evaluated at arbitrary grid points. The calculation of the spherical harmonic coefficients occurs in a second step by applying tailored quadrature formulas, as introduced in Section 6.2.1.

6.2.1 Quadrature Methods

Quadrature formulas in general, as described for example by BRONSTEIN and SEMENDJAJEW (1995), are a widely-used tool for the numerical evaluation of definite integrals. In the context of this thesis, quadrature procedures are applied to calculate spherical harmonic coefficients from discrete gravity field values provided on a spherical grid. This results in the numerical solution of the integral (3.39),

$$c_{nm} = \frac{1}{4\pi}\int_\Omega f(\mathbf{x})Y_{nm}(\mathbf{x})d\Omega. \tag{6.5}$$

The integral is to be approximated by a weighted sum of function values $f(\mathbf{x}_i)$ at given points $\mathbf{x}_i$ resulting in

$$c_{nm} = \frac{1}{4\pi}\sum_{i=1}^{I} w_i f(\mathbf{x}_i)Y_{nm}(\mathbf{x}_i), \tag{6.6}$$

where the weights w_i can be regarded as the surface element associated with the grid point. The weights sum up to the surface area of the unit sphere,

$$\sum_{i=1}^{I} w_i = 4\pi. \tag{6.7}$$

Eq. (6.6) only provides the correct result if the orthonormality of the spherical harmonics is preserved when converting the continuous integral to a discrete (weighted) sum, i.e. if the following relationship holds for the discrete case,

$$\langle Y_{nm}(\mathbf{x}), Y_{n'm'}(\mathbf{x})\rangle_W = \frac{1}{4\pi}\sum_{i=1}^{I} Y_{nm}(\mathbf{x}_i)Y_{n'm'}(\mathbf{x}_i)\, w_i = \delta_{nn'}\delta_{mm'}. \tag{6.8}$$

This is not generally true for arbitrary point distributions and corresponding weights w_i. The problem of discrete orthogonality can be addressed by investigating the corresponding weighted least squares adjustment procedures of estimating spherical harmonic coefficients from gridded data, as, for example, described by SNEEUW (1994). If the basis functions Y_{nm} are combined in the design matrix $\mathbf{A}$, and the weights are combined in the diagonal matrix $\mathbf{W}$, the least squares estimate of the coefficients $\hat{\mathbf{c}}$ is given by

$$\hat{\mathbf{c}} = (\mathbf{A}^T\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}\mathbf{f}. \tag{6.9}$$

In order to compare this result with a given quadrature formula, Eq. (6.6) can be formulated in matrix notation as well,

$$\mathbf{c} = \mathbf{A}^T \mathbf{W} \mathbf{f}. \tag{6.10}$$

Obviously, the two results are identical in case of a normal equation matrix $\mathbf{N} = \mathbf{A}^T\mathbf{W}\mathbf{A}$ being the identity matrix,

$$\mathbf{N} = \mathbf{A}^T \mathbf{W} \mathbf{A} = \mathbf{I}. \tag{6.11}$$

Every element of $\mathbf{N}$ represents the weighted inner product of the spherical harmonics according to Eq. (6.8). Therefore, the values of $\mathbf{N}$ apart from the main diagonal represent the amount of non-orthogonality of the corresponding discrete spherical harmonics. How much the normal equation matrix differs from a diagonal matrix depends on the grid point density on the one hand and on the grid pattern and the corresponding weights on the other hand. Therefore, it can be stated that a quadrature formula is the better suited for calculating spherical harmonic coefficients, the more accurately the orthonormality relations for spherical harmonics (3.32) are preserved.

In the following, several of these quadrature formulas will be introduced, each being characterized by the distribution of their quadrature nodes on the one hand and by the corresponding quadrature weights on the other hand. Firstly, two exact quadrature rules are investigated, namely the Gauss-Legendre quadrature and the Driscoll-Healy quadrature. Subsequently, the quadrature method is also tried out with the remaining point distributions described in Section 3.5.1.

6.2.1.1 Driscoll-Healy Quadrature

It was found out by Neumann (1838) that exact numerical quadrature up to a spherical harmonic degree $N = L - 1$ is possible by using $2L$ circles of latitude. They can be chosen arbitrarily, but have to be distinct. For example, an equidistant distribution of parallels is possible. This quadrature method is also quoted as 'Neumann's first method', see, for example, Sneeuw (1994). Driscoll and Healy (1994) have developed a closed representation for the weights of such an equi-angular spaced point distribution. These weights are applied in the following investigations, therefore the corresponding quadrature method is identified as Driscoll-Healy quadrature. The weights for the numerical integration along parallels are the analytical solution of

$$\sum_{j=0}^{2L-1} \bar{w}_i P_k(cos\frac{j\pi}{2L}) = 2\delta_{k0}. \tag{6.12}$$

This leads to the following weights, their derivation can be found in Driscoll and Healy (1994) with additional explanations provided in Mohlenkamp (1997),

$$\bar{w}_i = \frac{4}{2L}\sin(\vartheta_i)\sum_{l=0}^{L-1}\frac{\sin\left[(2l+1)\,\vartheta_i\right]}{2l+1}. \tag{6.13}$$

To obtain the weights to be applied in Eq. (6.6), the $\bar{w}_i$ have to be multiplied with $\Delta\lambda$,

$$w_i = \bar{w}_i \cdot \Delta\lambda. \tag{6.14}$$

6.2.1.2 Gauss-Legendre Quadrature

This method can be found in the work of Neumann (1838) as well, for this reason it is also referred to as 'Neumann's second method' (see, for instance, Sneeuw (1994) for a historical review on the different quadrature methods). Gaussian quadrature methods are widely applied in numerical analysis, here can be referred, e.g., to LANCZOS (1956), and are used in geodetic applications as well, see, for example, PAYNE (1971) and COLOMBO (1981). In contrast to the Driscoll-Healy quadrature described above, the Gauss-Legendre quadrature allows the recovery of a spherical harmonic expansion of degree $N = L - 1$ from only L circles of latitude. On the other hand, the parallels cannot be chosen arbitrarily, but have to be located along the zeros of the Legendre polynomial of degree L. Therefore, the quadrature nodes of the Gauss-Legendre quadrature method coincide with the grid points of the Gaussian grid, as described in Section 3.5.1.2. They feature equi-angular spacing along the circles of latitude, whereas along the meridians the nodal points are located at the L zeros of the Legendre polynomial of degree L. The orthogonality relations can best be identified when considering the specification of the spherical harmonic basis functions, as given by Eq. (3.47). It features a separation of the Y_{nm} into a trigonometric function depending only on the longitude λ and the associated Legendre functions depending only on the co-latitude ϑ. This allows to separately investigate the behavior of the basis functions in longitudinal and in latitudinal direction. The determination of the spherical harmonic coefficients can be split into a two-step procedure,

$$\left.\begin{array}{c} a_m(\vartheta) \\ b_m(\vartheta) \end{array}\right\} = \frac{1}{(1+\delta_{m0}\pi)} \int_0^{2\pi} f(\mathbf{x}) \left\{ \begin{array}{c} \cos(m\lambda) \\ \sin(m\lambda) \end{array} \right\} d\lambda, \tag{6.15}$$

$$\left.\begin{array}{c} c_{nm} \\ s_{nm} \end{array}\right\} = \frac{1+\delta_{m0}}{4} \int_0^{\pi} \left\{ \begin{array}{c} a_m(\vartheta) \\ b_m(\vartheta) \end{array} \right\} P_{nm}(\cos\vartheta) \sin\vartheta d\vartheta. \tag{6.16}$$

In the following, the analysis of the discretization of the integrals (6.15) and (6.16) will be described separately.

Longitudinal Direction The equi-angular spacing in longitudinal direction features $2L$ points along each circle of latitude. On such a regular configuration, the trigonometric functions obey discrete orthogonality relations,

$$\begin{aligned} &\sum_{i=0}^{2L-1} \cos m\lambda_i \cos \bar{m}\lambda_i = (1+\delta_{m0}+\delta_{mL})\, L\, \delta_{m\bar{m}}, \\ &\sum_{i=0}^{2L-1} \sin m\lambda_i \sin \bar{m}\lambda_i = (1+\delta_{m0}+\delta_{mL})\, L\, \delta_{m\bar{m}}, \\ &\sum_{i=0}^{2L-1} \cos m\lambda_i \sin \bar{m}\lambda_i = 0. \end{aligned} \tag{6.17}$$

Thus the conversion from the integral in Eq. (6.15) to a discrete sum delivers the exact result. The separate calculation in longitudinal direction is also beneficial from a computational point of view, as fast Fourier techniques (FFT) can be applied to the evaluation of Eq. (6.15).

Latitudinal Direction More crucial is the quadrature along the meridians. Here it can be made use of one-dimensional Gaussian quadrature procedures, as, for example, described in BRONSTEIN and SEMENDJAJEW (1995), applied to function values evaluated at points defined by $t_i = \cos\vartheta_i$. By the Gaussian quadrature rule a polynomial of degree $2L - 1$ can exactly be integrated by the evaluation of L function values,

$$\int_a^b f(t)dt = \sum_{i=1}^{L} f(t_i)w_i, \tag{6.18}$$

if the evaluation nodes t_i are chosen as the zeros of orthogonal polynomials. In case of the Gauss-Legendre quadrature, the orthogonal polynomials are the Legendre polynomials P_n, [1] and the integration interval equals $[a, b] = [-1, 1]$. The weights w_i can be derived using the characteristics of orthogonal polynomials,

---

[1] Equivalent Gaussian quadrature methods exist for Laguerre, Hermite, and Chebychev polynomials as well. They each require an additional, specific weighting function $q(t)$ to be introduced into the integral in (6.18). In case of the Gauss-Legendre quadrature, this weighting function equals a constant with the value of one and is, therefore, omitted in the above considerations.

as explained in more detail in Appendix B. For the one-dimensional quadrature along the meridians, the weights at a given co-latitude are ϑ_i depend on the number of parallels L and are given by

$$\bar{w}_i(L) = \frac{2}{(1-t_i^2)(P_L'(\cos(\vartheta_i)))^2}, \tag{6.19}$$

cf. STROUD and SECREST (1966). Again, the multiplication with $\Delta\lambda$ yields the final weights to be used in Eq. (6.6),

$$w_i = \bar{w}_i \cdot \Delta\lambda. \tag{6.20}$$

The relationship between the recoverable polynomial of degree $2L-1$ and the at maximum possible spherical harmonic degree $N = L-1$ to be determined from the given point distribution will be investigated in the following.

6.2.1.3 Maximum Degree Determinable by Exact Methods

The spherical harmonic analysis with the goal of calculating spherical harmonic coefficients from function values is formulated by Eq. (6.5). Considering Eq. (3.38), it becomes obvious that this requires the integration of products of spherical harmonics Y_{nm}, as illustrated by

$$c_{nm} = \frac{1}{4\pi}\int_\Omega f(\mathbf{x})Y_{nm}(\mathbf{x})d\Omega = \frac{1}{4\pi}\int_\Omega \left(\sum_{\bar{n}=0}^{N}\sum_{\bar{m}=-\bar{n}}^{\bar{n}} c_{\bar{n}\bar{m}}Y_{\bar{n}\bar{m}}(\mathbf{x})\right)Y_{nm}(\mathbf{x})d\Omega. \tag{6.21}$$

Thus the key point of the discretization process is to deliver a quadrature method valid for products of two spherical harmonics. Again, interpreting the integration in latitudinal and longitudinal direction separately leads to the following integrals,

$$\int_0^{2\pi}\int_0^{\pi} P_{nm}(\cos\vartheta)\left\{\begin{array}{c}\cos(m\lambda)\\ \sin(m\lambda)\end{array}\right\}P_{\bar{n}\bar{m}}(\cos\vartheta)\left\{\begin{array}{c}\cos(\bar{m}\lambda)\\ \sin(\bar{m}\lambda)\end{array}\right\}\sin\vartheta d\vartheta. \tag{6.22}$$

For $m = \bar{m}$ this results in evaluating the integral

$$\int_0^{\pi} P_{nm}(\cos\vartheta)P_{\bar{n}m}(\cos\vartheta)\sin\vartheta d\vartheta. \tag{6.23}$$

The product of the two associated Legendre functions $P_{nm}(\cos\vartheta)$ and $P_{\bar{n}m}(\cos\vartheta)$ is again a polynomial of degree at most $2N$ when N is the maximum degree to be determined. Therefore, the applied quadrature rule along the meridians needs to have an accuracy level (i.e. the at maximum determinable polynomial degree) of at least $2N$. If quadrature nodes are located arbitrarily, l nodes can provide an accuracy level of $l-1$ (BRONSTEIN and SEMENDJAJEW 1995). Therefore, at least $2N+1$ circles of latitude are required to achieve the accuracy level of $2N$ and thus to exactly evaluate the integral (6.23). This illustrates that, in case of the Driscoll-Healy quadrature, the $2L$ points along the meridians allow the determination of a maximum spherical harmonic degree $N = L-1$. If the quadrature nodes are chosen at the zeros of the Legendre polynomial of degree L, as in case of the Gauss-Legendre quadrature, the accuracy level of the corresponding quadrature rule can be enhanced. In this case l nodes allow an accuracy level of $2l+1$, or, in other words, the L meridians allow the recovery of polynomials up to degree $2L+1$. From this it results that the at maximum recoverable spherical harmonic coefficients are again of degree $N = L-1$. This can be shown, as the product of the Legendre functions has then a maximum degree of $2\cdot(L-1) = 2L-2$ which is smaller than the possible polynomial degree of $N = L-1$. In contrast to the Driscoll-Healy quadrature, however, this resolution can be achieved with only about half the number of circles of latitude. As both grids feature the same number of $2L$ grid points in longitudinal direction, the Gauss-Legendre quadrature requires only half the number of overall evaluation points. This is the reason why the Gauss-Legendre quadrature will be applied to the calculation of spherical harmonic expansions from regional gravity field solutions. The drawback of the irregular spacing of the quadrature nodes is not relevant, as the continuous representation of the regional solutions by radial basis functions allows the evaluation of the functionals at arbitrary points.

6.2.1.4 Quadrature in Case of Arbitrary Point Distributions

In principle, the quadrature formula (6.6) can be applied to any kind of point distribution. For reasons of completeness and in order to point out the differences to the exact quadrature procedures described above, the performance of the rest of the grids, described in Section 3.5.1, have been investigated as well. Thereby special emphasis is placed on the question of discrete orthogonality of the spherical harmonic basis functions, as this defines the accuracy of the quadrature method. In Fig. 6.1 the different normal equation matrices, as given by Eq. (6.11), are displayed for the calculation of spherical harmonic coefficients up to degree $n = 10$. The elements of the normal equation matrices contain the products of the spherical harmonic basis functions and, therefore, illustrate the discrete orthogonality relations.

Geographical Grid The quadrature weights are chosen as the surface area associated with each grid point,

$$w_i = \int\limits_{\lambda_i-\frac{\Delta\lambda}{2}}^{\lambda_i+\frac{\Delta\lambda}{2}} \int\limits_{\vartheta_i-\frac{\Delta\vartheta}{2}}^{\vartheta_i+\frac{\Delta\vartheta}{2}} = 2\cdot\Delta\lambda\sin(\Delta\vartheta)\sin(\vartheta_i). \tag{6.24}$$

The discrete orthogonality relations existing in case of the geographical grid can be investigated as to the longitudinal and latitudinal direction separately, as described above. Along the circles of latitude, the $2L$ equi-angular spaced points again preserve the discrete orthogonality of the trigonometric functions, as given by Eq. (6.17). Therefore, the C_{nm} and S_{nm} are orthogonal for different orders m, and, within the same order, the sine and cosine terms are orthogonal as well. Along meridians, the orthogonality among the associated Legendre functions is destroyed by the transition from the continuous to the discrete case. But the (anti-)symmetry characteristics of the associated Legendre functions,

$$P_{nm}(-\cos\vartheta) = (-1)^{n-m}P_{nm}(\cos\vartheta), \tag{6.25}$$

can be exploited. They ensure within the same order the independence of the coefficients of even degrees and the coefficients of odd degrees. The resulting normal equation matrix is displayed in the upper left part of Fig. 6.1. The coefficients are ordered by order, within each order by degree, with c_{nm} and s_{nm} alternating. The orthogonality between different orders, between sine and cosine, and between even and odd degrees becomes evident. The blocks along the main diagonal show the dependencies within the same order, the lack of further blocks indicates the orthogonality among different orders. Reordering the matrix by combining the sine and cosine terms and within each of those the even and odd degrees, leads to the well-known block-diagonal structure. The orthogonality relations and the resulting normal equation matrices for different numbering schemes in case of the geographical grid can be found in SCHUH (1996).

Reuter Grid Basically, the Reuter grid shows the same regularities and symmetries as described for the geographical grid. Along each circle of latitude, the spacing between grid points is again equi-angular, and the co-latitudes are located symmetrically with respect to the equator. Therefore, in principle, the same discrete orthogonalities are valid, as is the case for the geographical grid. This becomes evident in the upper right part of Fig. 6.1 by the blocks along the main diagonal exhibiting equal structure. Their magnitude is larger due to the declining point density towards the poles. The normal equation matrix of the Reuter grid does, however, possess additional correlation blocks with non-zero elements apart from those described above. They are caused by the circles of latitude near the poles for which the number of points γ_i calculated by Eq. (3.103) becomes very small. The resulting under-sampling of the trigonometric functions yields correlations between distinctive orders m and $\bar{m}$. As to the quadrature weights, the surface elements calculated from the Voronoi cells, as described in Section 3.5.2.1, can be applied. But due to the approximately uniform distribution, the use of these weights influences the quadrature results significantly less than, for example, in case of the geographical grid.

Triangle Vertex and Triangle Center Grid For the two different triangle grids, the normal equation matrices look very much alike. Therefore, only the one for the grid triangle vertex is displayed in the middle left part of Fig. 6.1. But the conclusions are valid for the triangle center grid as well. Even though the grids are not explicitly constructed based on equal- angular spacing between grid points, quite similar structures can be observed compared to the Reuter grid. The original icosahedron features a somehow regular and symmetrical structure. And obviously, certain symmetries and regular distributions enabling discrete orthogonalities are preserved during the densification. Again, the surface elements calculated from the Voronoi cells can be used as quadrature weights.

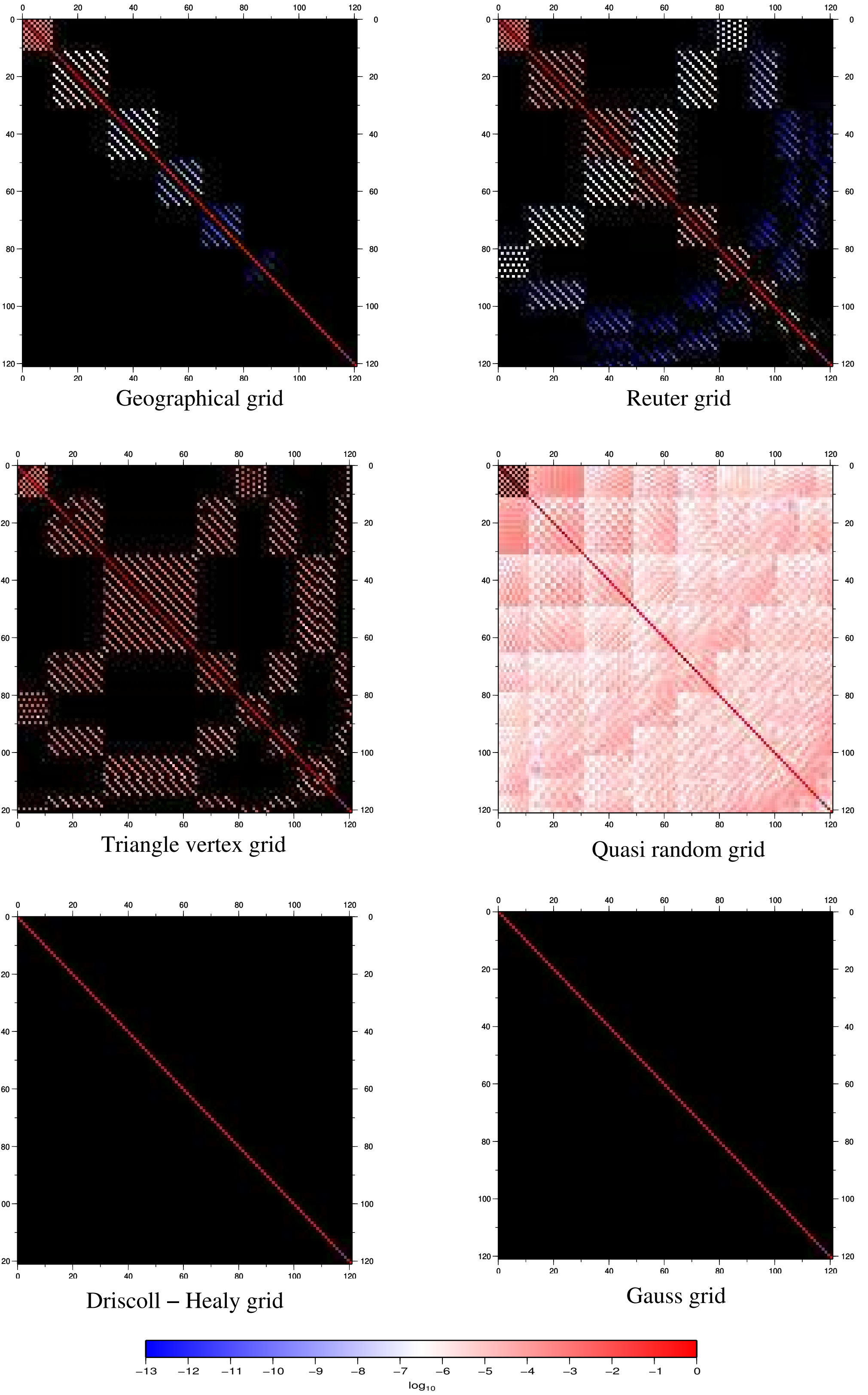
Quasi Random Grid In contrast to the other point distributions described in this chapter, the quasi random grid does not feature any discrete orthogonalities among different spherical harmonic basis functions, as illustrated by an almost fully occupied normal equation matrix in the middle right part of Fig. 6.1. The grid points are neither located regularly along parallels, nor are they arranged symmetrically to the equator. The only exception is the order $m = 0$, as here the orthogonality between odd and even degrees can again be observed. This can be explained by the fact that when the dependency on the longitude λ is omitted (as is the case for the spherical harmonics of order zero), the quasi random sequence positions the grid points symmetrically to the equator. Again, the surface elements calculated from the Voronoi cells can be used as quadrature weights.

6.2.1.5 Practical Calculations and Aliasing

Weight Matrix The quadrature weights, as determined in case of the Gauss-Legendre and the Driscoll-Healy quadrature, are chosen according to the requirement of realizing discrete orthogonality of the spherical harmonic basis functions. The weights can be combined in the weight matrix $\mathbf{W}$, which results in the normal equation matrix (6.11) becoming the identity matrix. This is only the case, however, for this particular choice of $\mathbf{W}$. Introducing an additional weight matrix, e.g. accounting for the stochasticity of the data, will destroy the orthogonality. So far, the patching process has been performed without error propagation from the regional to the global solutions. Therefore, no additional weight matrix is introduced. If the error propagation will be included in the process, this aspect will have to be taken into account.

Aliasing An aliasing effect occurs if a signal is undersampled, i.e. if the number of sampling points is not sufficient to completely reconstruct the original signal. The result is that the signal at higher frequencies become indistinguishable from that at certain lower frequencies (they are said to become 'aliases' of each other); thus they distort or create signal at lower frequencies. The aliasing effect is well-known from signal analysis, see, for example, Smith (1997). In order to investigate the corresponding effect in the case of spherical harmonics on the sphere, it proves to be reasonable to again investigate the behavior in longitudinal and latitudinal direction separately. In the following, the different effects are described, and each of them is illustrated by an example in Fig. 6.2. In the context of calculating spherical harmonic coefficients from gravity field functionals sampled at discrete data points, the aliasing problem has to be kept in mind. The grid (and thus the sampling) used for the (exact) quadrature methods described above has to be chosen sufficiently dense to avoid the effects that will be described in more detail below.

As pointed out in Section 6.2.1.3, the point setting of a Gauss-grid with L parallels allows the exact quadrature in latitudinal direction (i.e. along the meridians) up to a spherical harmonic degree of $N = L - 1$ using the Gauss-Legendre quadrature method. The required inner products of two associated Legendre functions of degree n and $\bar{n}$, respectively, are accurate in the discrete case as long as $n + \bar{n} < 2L$. As soon as the original signal contains frequencies higher than N, an aliasing error occurs. If $n = L + k$, then the inner products for $\bar{n} \geq L - k$ cannot be calculated correctly. The upper part of Fig. 6.2 illustrates this effect. Exemplarily, it shows the errors in spherical harmonic coefficients calculated from an original signal expanded up to degree $N = 80$. The quadrature was performed on a grid with $L = 60$ circles of latitude, which would have allowed the exact determination of coefficients up to degree $N = 59$. As n, in this scenario, can reach a degree up to $n = L + 20$, the coefficients of degree higher than $\bar{n} = L - 20 = 40$ are affected due to the



Geographical grid

Reuter grid

Triangle vertex grid

Quasi random grid

Driscoll – Healy grid

Gauss grid

Figure 6.1: Normal equation matrices for the calculation of spherical harmonic coefficients up to degree $N = 10$ from gridded data

undersampling. In contrast to Fourier analysis, there is no one-to-one correspondence between a distinctive high frequency component being mapped onto a single lower frequency (as will be described below in case of the longitudinal direction). On the contrary, the undersampled function appears as a linear combination of the lower frequency functions up to degree $\bar{n} \geq L - k$. Only in case of $n + \bar{n}$ odd, the odd symmetry characteristics of the Legendre functions provides for the inner product to become correctly zero.

The quadrature in longitudinal direction corresponds to the sampling of trigonometric functions; therefore aliasing phenomena are equal to those arising in the case of Fourier analysis. In this context, the so-called sampling theorem declares that in order to reconstruct a signal exactly, the sampling frequency f_s has to be more than twice the highest frequency f_{max} present in the original signals,

$$f_s > 2f_{max}. \tag{6.26}$$

The frequency representing half of the sampling rate is denoted as Nyquist frequency, and all frequencies smaller than the Nyquist frequency can be reconstructed from the present sampling. As soon as higher frequencies are present, the aliasing effects occur. In case of the Gauss grid, there are $2L$ sampling points along each circle of latitude, thus the Nyquist frequency equals a spherical harmonic order of $m = L$, and orders up to $m = L - 1$ can be reconstructed correctly. Concerning the aliasing phenomena, it can be differentiated between reflective aliasing and periodic aliasing. The term 'reflective aliasing' specifies the fact that a frequency of order $m = L + k$ corrupts the coefficients of order $L - k$, as can be understood by considering

$$\begin{aligned} \cos\left((L+k)\, j\, \frac{2\pi}{2L}\right) &= \cos\left((2L + (k-L))\, j\, \frac{2\pi}{2L}\right) \\ &= \cos(j\, 2\pi)\, \cos\left((k-L)\, j\, \frac{2\pi}{2L}\right) - \sin(j\, 2\pi)\, \sin\left((k-L)\, j\, \frac{2\pi}{2L}\right) = \cos\left((L-k)\, j\, \frac{2\pi}{2L}\right) \end{aligned} \tag{6.27}$$

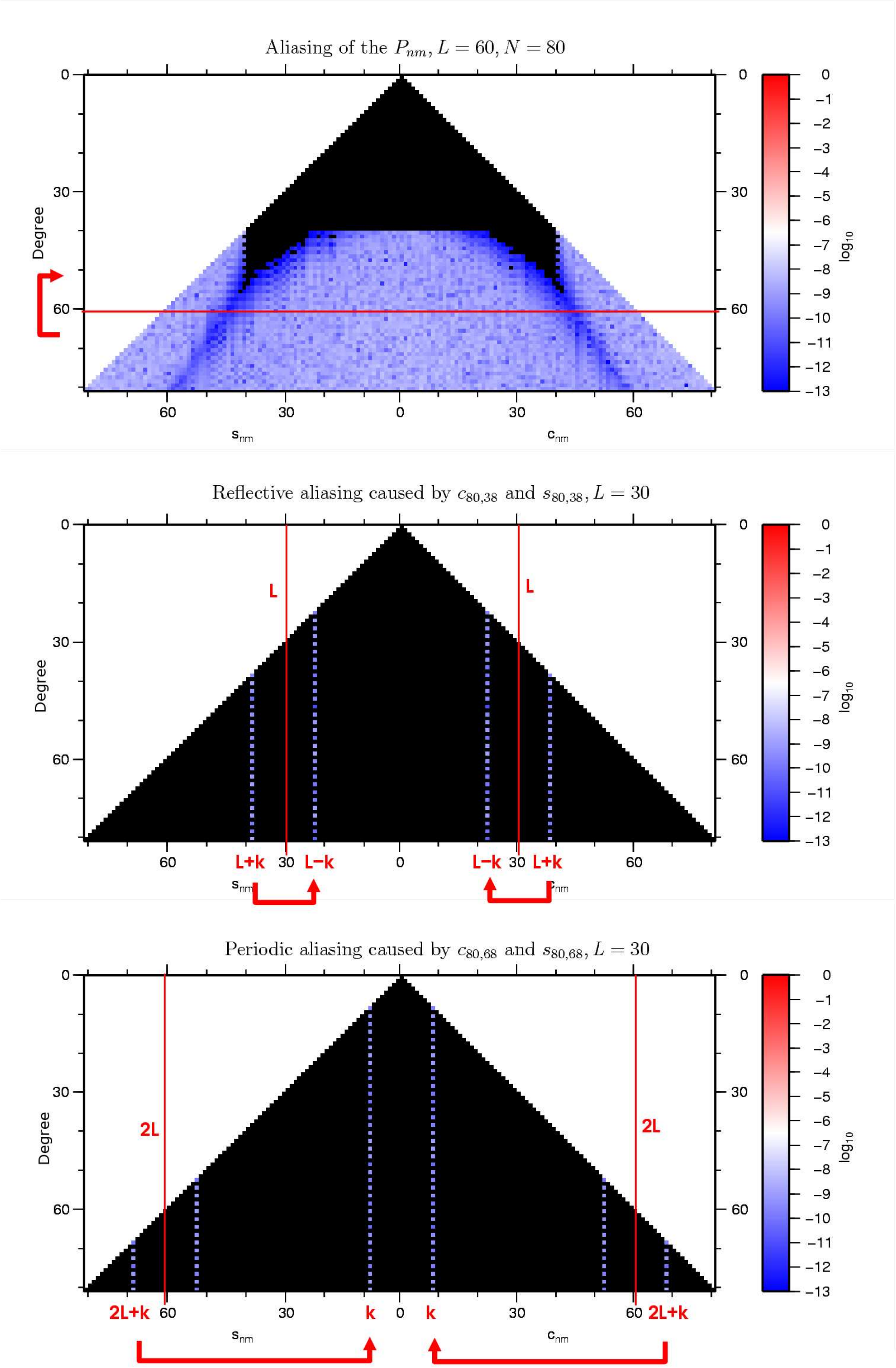
for $j = 0 \ldots 2L$. Equivalent considerations can be made concerning the sine functions resulting in the relation

$$\sin\left((L+k)\, j\, \frac{2\pi}{2L}\right) = -\sin\left((L-k)\, j\, \frac{2\pi}{2L}\right). \tag{6.28}$$

This is illustrated in the middle part of Fig. 6.2 for a point setting of $L = 30$ parallels enabling the reconstruction up to a maximum order of $m = 29$. The original signal that was to be reconstructed only consisted of the coefficients $c_{80,38}$ and $s_{80,38}$, all the other coefficients being set to zero. The results confirm that order $m = L + 8 = 38$ is mapped onto order $\bar{m} = L - 8 = 22$. The fact that the single coefficients influence coefficients of different degrees again demonstrates the aliasing effect in latitudinal direction, as described above. Apart from the reflective aliasing, the periodic character of the trigonometric functions leads to periodic aliasing as well. This can be understood from

$$\exp\left(i(2L+k)\, j\, \frac{2\pi}{2L}\right) = \exp(ij2\pi)\, \exp\left(ikj\, \frac{2\pi}{2L}\right) = \exp\left(ikj\, \frac{2\pi}{2L}\right). \tag{6.29}$$

This shows that the functions $\exp(i(2L + k)\lambda)$ are indistinguishable from the function $\exp(ik\lambda)$, thus coefficients of order $m = 2L + k$ cause errors in the coefficients of order $m = k$. This effect is displayed in the lower part of Fig. 6.2, again for a setting with $L = 30$. This time all coefficients apart from $c_{80,68}$ and $s_{80,68}$ were set equal to zero. The periodic aliasing can be observed by the order $m = 68 = 2L + k = 60 + 8k$ causing corruption of the order $\bar{m} = 8$. At the same time order $\bar{m} = 2L - k = 52$ is corrupted as well, as the reflective aliasing described above also occurs for multiples of L.


Aliasing of the $P_{nm}, L = 60, N = 80$

Reflective aliasing caused by $c_{80,38}$ and $s_{80,38}, L = 30$

Periodic aliasing caused by $c_{80,68}$ and $s_{80,68}, L = 30$

Figure 6.2: Aliasing effects caused by the undersampling of spherical harmonic functions. Upper part: aliasing in latitudinal direction, middle part: reflective aliasing in longitudinal direction, lower part: periodic aliasing in longitudinal direction

7. Calculations and Results

In the following chapter, the practical calculations in the context of regional gravity field refinement are described. The chapter starts with the introduction of the programming system GROOPS that was designed and applied in the calculations. In the second section, a simulation study is presented that is to investigate the practical implications of theoretical convergence issues discussed in the above chapters of this thesis. Subsequently, the regionally refined gravity field solutions calculated from the data of the new satellite missions are described. In case of CHAMP and GRACE, this implies the analysis of real data. In case of GOCE, a simulation scenario is presented based on a combined analysis of (simulated) GRACE and GOCE data.

7.1 (Real) Data Analysis with the Programming System GROOPS

The following section is dedicated to the description of the gravity field analysis software GROOPS (**GR**avity field **O**bject **O**riented **P**rogramming **S**ystem). It has been developed and implemented by our group during the last six years, and it comprises, among others, all the concepts and calculation procedures described and applied within this thesis. GROOPS is implemented in C++ and features an object-oriented design implying a completely modularized structure. This evokes a high degree of flexibility to deal with the various tasks required in the context of a gravity field recovery process.

The following introduction will be limited to those parts of the software system that are directly used in the calculation of static (regional) gravity field solutions. A simplified schematic overview of these constituents can be found in Fig. 7.1. It displays the dependencies between the most important classes (indicated by blue boxes) and programs (identified by yellow boxes). In each of the classes, the interface of the class is pointed out in the blue box, the inherited classes that implement the interface are listed in the respective box below. The arrows in the diagram represent dependencies between classes and programs. An arrow pointing from one class to another indicates that the former is required by the latter (even though this does not always have to be true for each one of the inherited classes). Arrows from programs to classes, or vice versa, specify that the class is needed as input for the program, or that the output of the program is used by the given class, respectively. (For reasons of clarity, some less important arrows have been neglected in the diagram.) In the following, the different classes and programs (printed in bold face) and their respective interactions will be explained.

The analysis procedure starts with the applied data sets; the system can process orbit data, K-band observations, gradiometer data as well as the information provided by accelerometer and star camera sensors. The data can either be obtained in form of real satellite data or in form of simulated data. When dealing with the analysis of real data, the first step requires the import of the various data sets and the conversion into a format processible by the analysis software. In case of a simulation scenario, GROOPS enables a consistent closed-loop analysis. An extensive simulation tool within the programming system allows the system not only to be used for the processing of the gravity data, but also to be applied to the generation of simulated data sets. A tailored noise model implemented for each of the observation types allows to create simulation scenarios that are as realistic as possible. At the beginning of the processing chain, the different data sets are imported into the program **ArcDesigner**. It is responsible for cutting out satellite data over specific regions (in case of a regional analysis) and for splitting the orbit into smaller pieces to enable the use of short arcs in global analysis, too. In order to determine whether a given satellite position is within a geographical region, the program needs the boundary of the region on the surface of the Earth. This information can be obtained from classes inherited from the interface **border** that define the limits of the regions. The areas can either be chosen as defined between two meridians and two circles of latitude (**rectangle**), in the shape of a spherical cap (**cap**), or by an arbitrary polygon of points on the sphere (**polygon**). Furthermore, Earth rotation information is required to transform the position into the Earth-fixed coordinate system; this is necessary to associate the satellite data (given in a quasi-inertial system) with positions on the Earth. The corresponding information is provided by elements of **earthRotation**, here can be chosen between the specifications given by the IERS (**iers1996** or **iers2003**) and simplified versions comprising only the rotation around the z-axis (**gmst** and **zAxis**), the latter two differing concerning their rotational angles. The resulting satellite data prepared by **ArcDesigner** serve as input in the set-up of the observation equations, as will be explained below. In addition to the satellite measurements, the disturbing forces acting on the spacecraft have to be considered. A short summary of the different disturbing forces and the applied models is given in Section 7.1.1. These models can either directly be included in the observation equations, or alternatively they can be calculated beforehand at the satellite's positions. This is performed by the program **ArcGravity** which calculates a reference acceleration including a reference gravity field solution (provided by elements of the class **referenceField**), tidal models (given by **tides**), and the de-aliasing product (also implemented as **referenceField**) to account for short-periodic changes of the atmosphere and the reaction of the ocean to these changes.

In the class **observation**, the observation equations are set up. GROOPS enables the selection of observation equations for precise orbit determination, for satellite-to-satellite tracking in the low-low mode, for gradiometer observations, and for terrestrial data sets. The classes **podIntegral**, **graceRange**, and **sgg** represent the observation equations described in the Sections 4.3.1, 4.3.2, and 4.3.3 and applied in the calculations in this thesis, as described in Section 7.3. The classes **podAcceleration**, **podEnergy**, and **graceEnergy** contain alternative functional models based on the acceleration approach and energy balance approach (both shortly described in Section 2.2.1) that have been implemented for reasons of comparison. The opportunity to directly use point values on the sphere as observations is offered by the class **terrestrial**. As input, the elements of **observation** require the observations themselves (preprocessed by **ArcDesigner**), Earth rotation information (given by **earthRotation**), and either the output of the program **ArcGravity** or the applied reference field and the models of the disturbing forces if they have not been calculated to the satellite's orbit beforehand. Furthermore, classes of the type **observation** have to know the desired modeling of the gravity field solution in terms of a set of unknown parameters. This is provided by classes inherited from the interface **representation**; here can be distinguished between **sphericalHarmonics** and **splines**. The additionally listed choices **timeLinear** and **timeFourier** refer to the modeling of time variabilities and will not be discussed in detail here. The interaction of the two classes **observation** and **representation** very well emphasizes the flexibility of the analysis software. The observation type and the gravity field representation can be chosen independently, and each one of the observation methods can either be used with a global representation in terms of spherical harmonics or with a regional modeling in terms of spline functions. In case of a modeling with radial basis functions, the class **representation** additionally requires the shape of the basis functions and their location on a spherical grid. The former is provided by the class **coefficients** inherited from the interface **kernel**, the shape coefficients can thereby be calculated by the program **KernelDegreevariance** from a given gravity field model. The nodal point setting is provided by **grid**, here all the point distributions described in Section 3.5.1 are implemented.

The class **observation** is required by the program **NormalsBuild** which then accumulates the normal equations. If desired, a regularization matrix can be included into the calculations, the respective matrix is calculated by different specifications of the program **Regularization**. These are in detail: **RegularizationDegreevariances** which applies the reciprocal degree variances of a given spherical harmonic model on the main diagonal of the regularization matrix, **RegularizationKaula** in which the degree variances are approximated by Kaula's rule of thumb, and **RegularizationSigma** which uses the reciprocal formal accuracies of the spherical harmonic coefficients on the main diagonal. For the regional adaption of the regularization process, the program **RegularizationBorders** is of particular importance, as it generates the regularization matrices for different regularization areas, as described in Section 5.2.5. In order to assign different basis functions to different regularization areas, the program requires an element of the class **border**. The program **RegularizationSplines** creates the fully occupied regularization matrix containing the inner products of splines as basis functions, as described in Section 5.2.3. The next step is the solving of the system of normal equations; different versions of a solver are implemented. A direct solution of the system of equations via Cholesky decomposition is possible using **SolverDirect**. The program **SolverVarianceComponent** iteratively adjusts the weighting factors for different normal equations (and the regularization matrix) using the variance component estimation procedure described in Section 5.2.4. An iterative solution of the system of equations, applying the method of preconditioned conjugate gradients, is given by **Solver-**

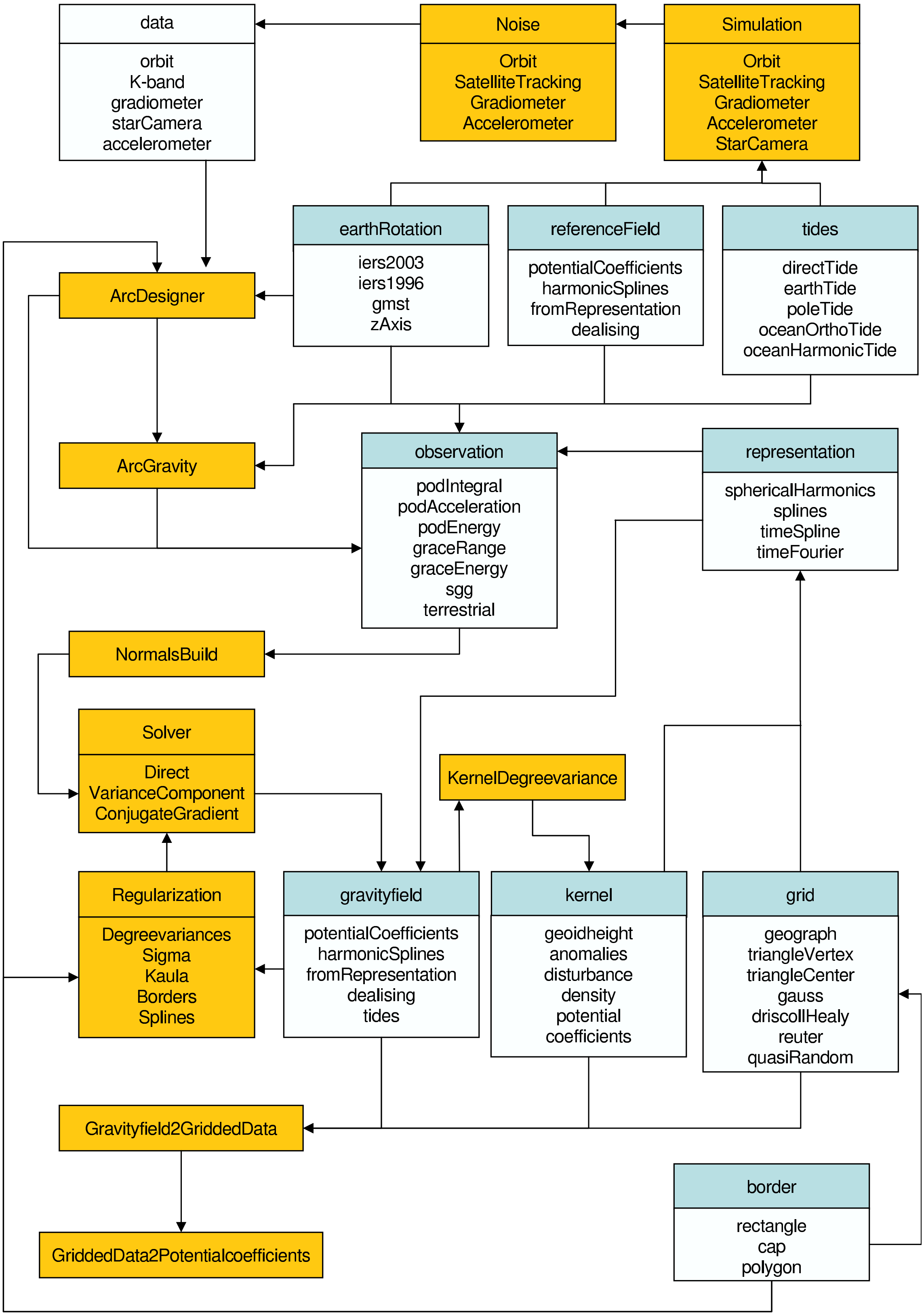
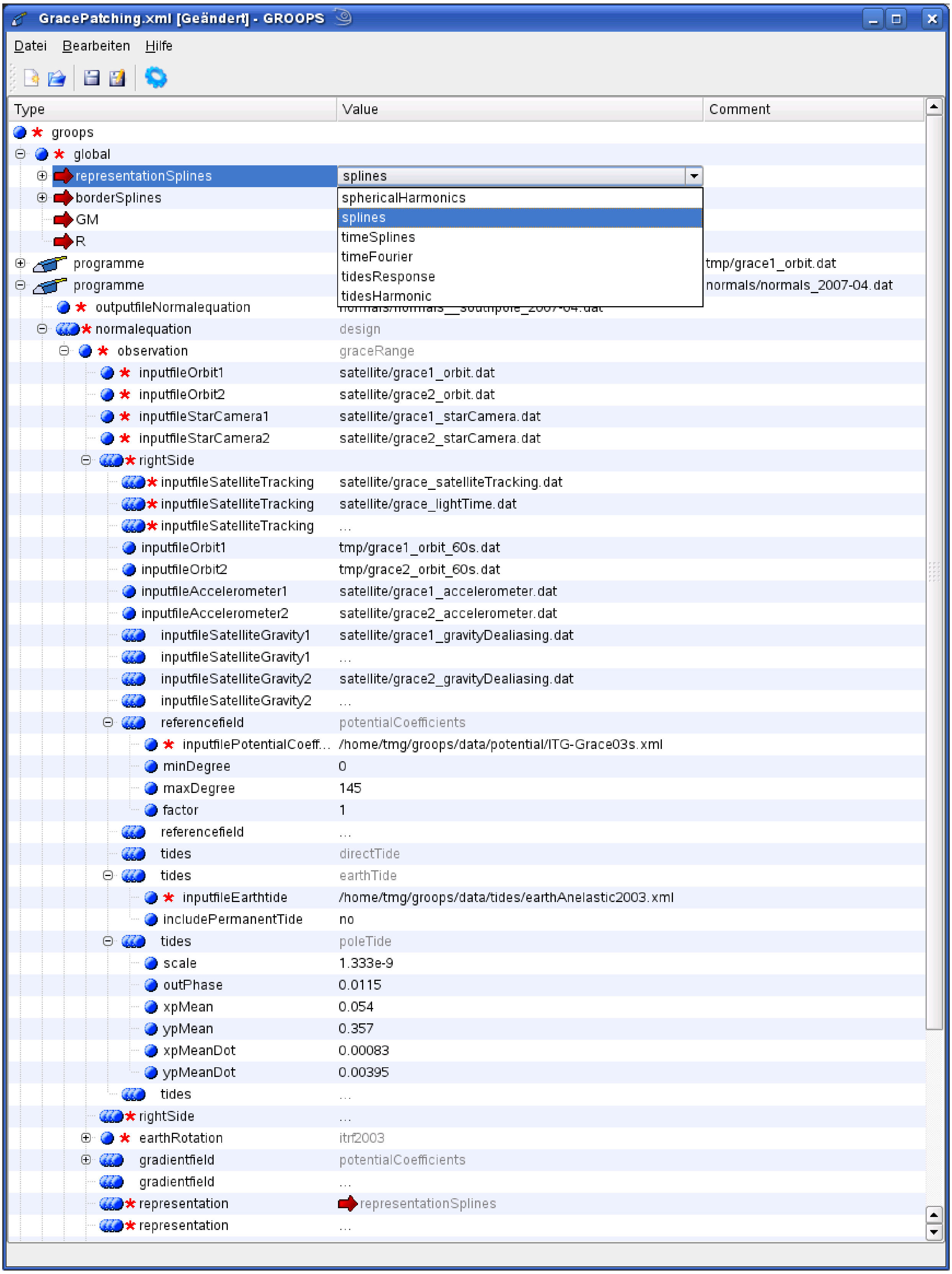


Figure 7.1: Modularized layout of the gravity field analysis software GROOPS showing the major classes (blue) and programs (yellow)



Type	Value	Comment
groops		
global		
representationSplines	splines	
borderSplines		
GM		
R		
programme		tmp/grace1_orbit.dat
programme		normals/normals_2007-04.dat
outputfileNormalequation	normals/normals__southpole_2007-04.dat	
normalequation	design	
observation	graceRange	
inputfileOrbit1	satellite/grace1_orbit.dat	
inputfileOrbit2	satellite/grace2_orbit.dat	
inputfileStarCamera1	satellite/grace1_starCamera.dat	
inputfileStarCamera2	satellite/grace2_starCamera.dat	
rightSide		
inputfileSatelliteTracking	satellite/grace_satelliteTracking.dat	
inputfileSatelliteTracking	satellite/grace_lightTime.dat	
inputfileSatelliteTracking	...	
inputfileOrbit1	tmp/grace1_orbit_60s.dat	
inputfileOrbit2	tmp/grace2_orbit_60s.dat	
inputfileAccelerometer1	satellite/grace1_accelerometer.dat	
inputfileAccelerometer2	satellite/grace2_accelerometer.dat	
inputfileSatelliteGravity1	satellite/grace1_gravityDealiasing.dat	
inputfileSatelliteGravity1	...	
inputfileSatelliteGravity2	satellite/grace2_gravityDealiasing.dat	
inputfileSatelliteGravity2	...	
referencefield	potentialCoefficients	
inputfilePotentialCoeff...	/home/tmg/groops/data/potential/ITG-Grace03s.xml	
minDegree	0	
maxDegree	145	
factor	1	
referencefield	...	
tides	directTide	
tides	earthTide	
inputfileEarthtide	/home/tmg/groops/data/tides/earthAnelastic2003.xml	
includePermanentTide	no	
tides	poleTide	
scale	1.333e-9	
outPhase	0.0115	
xpMean	0.054	
ypMean	0.357	
xpMeanDot	0.00083	
ypMeanDot	0.00395	
tides	...	
rightSide	...	
earthRotation	itrf2003	
gradientfield	potentialCoefficients	
gradientfield	...	
representation	representationSplines	
representation	...	

Figure 7.2: Screenshot of the graphic user interface of GROOPS

ConjugateGradient; it is especially useful in case of large systems of equations and, therefore, less essential when dealing with a regional recovery process.

The output made available by the different solver programs is a set of unknown parameters together with their formal accuracies. The solution derived for the unknown parameters can then be interpreted by the class **fromRepresentation** of type **gravityfield** which defines a gravity field from a parameter vector associated with a given representation. Further options of the interface **gravityfield** are the classes **potentialCoefficients** that constitute a gravity field from a given set of spherical harmonic coefficients, **harmonicSplines** creating a gravity field from a file containing spline parameters together with the respective spline kernel and the nodal point pattern, **dealiasing** representing the specifications provided by a de-aliasing product, and **tides** interpreting a given tide model. To evaluate and display the obtained gravity field solution, the program **GravityField2GriddedData** can be used to calculate gravity field functionals at given grid points. These grid points are again defined by the class **grid**. The classes of the interface **kernel** can be used to obtain the different functionals. These are in detail: the gravitational potential (**potential**, using the Poisson kernel), gravity disturbances (**disturbance**, Hotine kernel), geoid heights (**geoidheight**, applying the Poisson kernel divided by normal gravity), gravity anomalies (**anomalies**, Stokes kernel), and the density of a single layer (**density**).

As output, **GravityField2GriddedData** delivers a file with spherical coordinates of the grid points together with the corresponding gravity field values and the area weights associated with each grid point, as defined in Section 6.2.1. This output can, e.g., be applied in the graphic presentation of the respective gravity field. Furthermore, the program specifies statistical information about the gravity field such as a weighted RMS of the given solution. If individual regional patches with global coverage are supposed to be merged to obtain a global spherical harmonic solution, this can be carried out by the program **GriddedData2Potentialcoefficients**. It uses the file provided by **GravityField2GriddedData**, extracts the grid points and weights, and calculates the spherical harmonic coefficients by the corresponding quadrature rule.

The modularized design of GROOPS does not only allow a very flexible combination of the various elements, but it also helps to avoid the multiple implementation of items that are required at different stages of the analysis process. For example, the class **border** can be used for the clipping of the satellite data, for the definition of regionally adapted regularization matrices, and to calculate gravity field functionals for certain areas. Another example is the fact that an element of **kernel** is needed for the design of the shape of space localizing basis functions as well as for the conversion between different gravity field functionals, even though these two tasks might not seem to have too much in common at first sight.

The configuration of GROOPS is controlled using the Extensible Markup Language (XML). The adjustment of the individual programs and classes can either be performed by adapting the respective XML file, more comfortable is the use of the tailored graphic user interface. An exemplary screenshot of this interface is displayed in Fig. 7.2. It features pop-up menus for the individual programs and classes. The given example shows an extract of a regional gravity field determination from GRACE data using the integral approach. The program **NormalsBuild** is expanded, showing, e.g., the class **graceRange** of the interface **observation**. Required items are denoted by a red star, additional items are optional and depend on the respective application. The user interface offers the possibility of defining repeatedly used variables as global variables, these are indicated by a red arrow.

GROOPS exists in a version that can be run on an individual computer and in a parallelized version to be used on a multi-machine computing cluster.

7.1.1 Background Models

If the gravitational field of the Earth is to be determined from satellite observations, all the other forces acting on the satellite have to be reduced beforehand. These disturbing forces can either be approximated by appropriate models or have to be measured by the on-board accelerometer. A short summary of the background models that have to be accounted for during the calculations dealing with the current satellite missions is given in the following.

Direct Tides Apart from the gravitational force of the Earth, the satellite's orbit is influenced by the sun, the moon, and the planets in form of tidal forces. These tidal forces are the differences between the gravitational forces of the third bodies acting on the Earth on the one hand and on the satellite on the other hand. They can be calculated from the positions of the third bodies, listed in the Ephemerides DE 405 provided by the JPL (STANDISH 1998).

Solid Earth Tides The tidal forces described above cause a deformation of the Earth. This leads to mass displacements resulting in a change of the gravitational potential denoted as tidal potential of the solid Earth. The corresponding specifications are given in the IERS 2003 conventions (MCCARTHY and PETIT 2004).

Pole Tides The polar tides are induced by the centrifugal force of the polar motion which generates a deformation of the Earth and thus a change in the gravitational potential. The related specifications can again be obtained from the IERS 2003 conventions.

Ocean Tides The modeling of the oceanic reaction to the tidal forces and the resulting gravitational effects are determined using the ocean tide model FES2004 (LE PROVOST 2001).

De-aliasing Short periodic variations of the atmospheric masses and the reaction of the ocean to these mass variations have to be modeled and reduced, as otherwise they would distort the solution by aliasing effects. The models to account for the short periodic variations are, therefore, denoted as de-aliasing data (AOD1B), as specified in FLECHTNER (2005).

Non-gravitational Forces The non-gravitational forces acting on the satellite have to be separated from the gravitational forces. They cannot easily be approximated by models, therefore they have to be directly measured by the on-board accelerometer located in the satellite's center of mass. The non-gravitational forces are in detail the surface forces induced by atmospheric drag, solar radiation, and Earth albedo, as well as accelerations generated by steering thrusters.

Relativistic Correction The large velocities in combination with the high measurement accuracies would demand a relativistic formulation of the equation of motion. But since the differences are rather small, the effect can be compensated by relativistic corrections, as specified in the IERS 2003 conventions.

7.2 Simulation Study: Basis Functions

The following studies are performed in order to investigate the practical impact of the theoretical considerations outlined in the Sections 3.3.4 and 5.2.3. There the matter of functions belonging to the reproducing kernel Hilbert space of certain kernels was addressed. When specifying the space localizing spline functions Φ, it was argued that in a Hilbert space $\mathcal{H}_C$ defined by the covariance function (3.58) of the gravitational potential as reproducing kernel two basis functions are decorrelated with respect to their inner product, but possess an infinite norm, as shown by Eq. (5.31). The problem of infinite norm could be overcome by introducing a modification factor according to (5.33), $f_{mod} = \left(1/\sqrt{1+\epsilon}\right)^n$, leading to the modified spline kernels $\bar{\Phi}$. This causes the loss of the orthogonality as a drawback. In practical calculations, the inner product of the basis functions occurs in the construction of the regularization matrix as defined in Eq.(5.28), which implies that the elements of the regularization matrix are composed of these inner products. In case of bandlimited basis functions, naturally, no infinite norm can occur even without application of any modification factor. But again the regularization matrix cannot strictly be regarded as diagonal matrix, as the decorrelation of the spline functions with respect to the inner product $\langle \cdot, \cdot \rangle_C$ only exists if they are expanded up to $N = \infty$. Exemplarily, such a fully occupied regularization matrix $\mathbf{R}$ containing the inner products of bandlimited spline functions is displayed on the left side of Fig. 5.1. Nevertheless, the modification factor f_{mod} can also be applied to bandlimited basis functions, leading to matrices related to the one shown on the right side of Fig. 5.1. Despite these issues, it would provide advantages to consider the basis functions as uncorrelated and thus to approximate the regularization matrix by a diagonal (unit) matrix. Here especially the possibility of a regionally adapted regularization (see section 5.2.5) has to be mentioned, as this procedure necessarily requires that the regularization matrix is separable into different matrices belonging to different regularization regions. Such a separation would be impossible in case of a fully occupied matrix. Apart from the fact that the basis functions do not belong to the RKHS defined by $\mathcal{C}$, it can also be argued that the gravitational potential itself does not belong to the RKHS defined by its covariance function. This was illustrated in Eq. (3.68). There it was proposed (in accordance with Moritz 1980) that a modification of the covariance functions by a factor $f_{Cov} = (1+\epsilon)^n$, as given in Eq. (3.69), would take care of this problem, leading to the modified version $\bar{\mathcal{C}}$. According to (5.42), this would lead to the same regularization matrix as described for the modified spline kernels.

In the following, it will be investigated in how far the above considerations, and especially the approximation of the regularization matrix as a unit matrix, has any impact on practical calculations. Therefore, a simulation example was set up, using a CHAMP-like setting to calculate gravity field parameters from simulated orbit data. As pseudo-real field served the EGM96 (Lemoine et al. 1998) up to degree $n = 300$; the satellite positions were corrupted by simulated noise with 3 cm position accuracy. From this setting, a global solution was determined parameterized by spline functions. For this study, the calculation of a global spline solution instead of a regional one was regarded as suggestive to prevent the result from possible other artefacts such as truncation effects. Beforehand, it was investigated by calculating a global spherical harmonic solution that from this simulation setting gravity field features up to a resolution of approximately $N = 45$ can be recovered. In order to fully exploit this signal content, the spline functions were located on a triangle vertex grid of level 16 (leading to 2892 nodal points, approximately corresponding to a spherical harmonic expansion of $N = 50$ with 2601 parameters). The expansion of the basis functions themselves was also truncated at $N = 50$. Since a global spline solution can be transformed into a spherical harmonic expansion according to Eq. (6.4), the results can be evaluated in the frequency domain in terms of degree variances. The signal degree variances of different choices for basis functions and regularization matrices are displayed, as well as for each choice the difference degree variances compared to the pseudo-real field EGM96.

Scenario 1: Kernel Without Modification As a reference, a solution was calculated using the original spline kernel Φ and approximating the regularization matrix by the unit matrix $\mathbf{R} = \mathbf{I}$. In Fig. 7.3 to Fig.7.6, the resulting signal is illustrated by the orange line, and the differences compared to the EGM96 are represented by the red line. In a second step, the same solution was calculated with the only difference that the unit matrix was substituted by the fully occupied regularization matrix, resulting in the light blue signal and the dark blue error curve in Fig. 7.3. The differences between both solutions are marked by the dashed blue line. Since the lines are very close together, the crucial higher frequency part is magnified in each of the figures. But even in the enlarged version, the error curves of both solutions are almost indistinguishable, which can be recognized by the differences of the two solutions being by a factor 5-20 smaller than the errors. This leads to the conclusion that the application of the fully occupied regularization matrix does not provide any significant improvements, and that it can be approximated by the unit matrix without significant loss of accuracy when dealing with the original bandlimited spline functions.

Scenario 2: Modified Kernel, Unit Matrix In the second scenario (Fig. 7.4), the spline functions were modified by the modification factor described above. Two different factors ($f_{mod} = \sqrt{0.95}$ and $f_{mod} = \sqrt{0.97}$) were tested, and the results are illustrated by the light blue and dark green line, respectively. The error curves of both results are given by the corresponding dark blue and light green lines. In this case, the regularization matrix was again approximated by the unit matrix $\mathbf{R} = \mathbf{I}$. The modification dampens the high frequencies in the spline kernels, which obviously leads to a dampening of the signal as well, since the dark blue and light blue line run significantly below the orange line. This loss of signal in the high frequencies is also reflected by a very slight increase in the errors, making the light green and dark blue curve run marginally above the red line. Thus it can be assumed that when dealing with the modified spline functions, the use of the unit matrix in the regularization process leads to a loss in resolution and is, therefore, not advisable.

Scenario 3: Covariance Matrix Modified As mentioned above, not only the basis functions do not belong to the Hilbert space defined by the covariance function as reproducing kernel, but also the gravitational potential itself has an infinite norm within this space and, therefore, does not suffice the smoothness condition inflicted by the kernel. That is why the idea can arise to introduce the modified covariance function $\bar{\mathcal{C}}$ being slightly rougher than the original one. This results in the same regularization matrix as will be applied in Scenario 4, but here only the regularization matrix is modified and not the spline kernel itself. The results can be seen in Fig. 7.5. As expected, the less strict smoothness condition causes the signal in the higher frequencies to be less dampened compared to the original solution, as proved by the dark green ($f_{Cov} = 0.97$) and light blue line ($f_{Cov} = 0.95$) indicating the signal curves to run closer to the signal of the pseudo-real field EGM96. On the other hand, however, the errors represented by the light green and dark blue curves have become larger as well, giving evidence that this kind of modification of the covariance function does not increase the accuracy of the solution.

Scenario 4: Modified Kernel With Fully Occupied Covariance Matrix In the last scenario, the modified spline kernels $\bar{\Phi}$ mentioned above were used for the parameterization of the solutions, and the corresponding fully occupied regularization matrix was applied. The results are displayed in Fig. 7.6. The modification by a factor of $f_{mod} = \sqrt{0.95}$ leads to an amplification of the higher frequencies and an increase of the errors compared to the original solution. In case of a lesser modification (here shown for a factor of $f_{mod} = \sqrt{0.97}$, but the same is valid for arbitrary higher factors), the signal curve as well as the error curves almost coincide with that of the original solution. Obviously, the dampening due to the modification of the spline functions, as observed in the second scenario, is compensated if the appropriate regularization matrix is applied. However, the results do not show any significant improvement when compared to the original solution.

Summarizing, it can be stated that the modification of either the spline kernel or the covariance function have a dampening or amplifying effect on the higher frequency part of the solution, each leading to a worsening of the errors. When the modification of the spline kernels is met with the appropriate regularization matrix, the effect can be compensated. But in none of the different scenarios, a result was derived that is significantly better than the original solution calculated from the unmodified spline kernel with the unit matrix as approximation of the regularization matrix. Therefore, it can be concluded that the considerations described in Sections 3.3.4 and 5.2.3 do have a theoretical value when analyzing the behavior of space localizing basis functions from a mathematical point of view, but they do not seem to have any significant impact on the practical calculations. The above simulation study was only performed in detail for a CHAMP-like scenario, but tests have shown that also in case of GRACE and GOCE no significant drawback seems to be caused by the approximations described. Possibly, in case of future measurement concepts providing substantial increase in accuracy and resolution, the discussed issues can be expected to make a difference. But when dealing with the observations provided by the present satellite missions, as processed within the scope of this thesis, the specified approximations are definitively sufficient. This is especially valid when it is taken into account that the gain in accuracy achieved by the regionally adapted regularization process does considerably prevail any inaccuracies caused by the approximations made.

7.3 Gravity Field Solutions

In the following section, the regional gravity field recovery approach will be applied to the data of the satellite missions CHAMP, GRACE, and GOCE. In case of CHAMP and GRACE, the presented solutions are derived from the analysis of real data, in case of GOCE, a simulation scenario is generated. In principle, the procedure applied to the different data sets can be characterized by the following steps: At first, a global reference field is necessary, because the regional solutions are always calculated as refinements to a global solution. The reference field is usually parameterized by spherical harmonics, even though different parameterizations are possible as well. The global field is particularly responsible for covering the longer wavelength part of the gravity field spectrum; therefore, it is sufficient to be modeled up to a moderate spherical harmonic degree. The global solution is subtracted from the data, and in a second step the residual field is calculated in regional areas to derive regional refinements to the global reference field. To meet the regional character of these solutions, they are represented by space localizing radial basis functions. If only a regional gravity field is desired, the procedure is completed at this point. If one is interested in a global

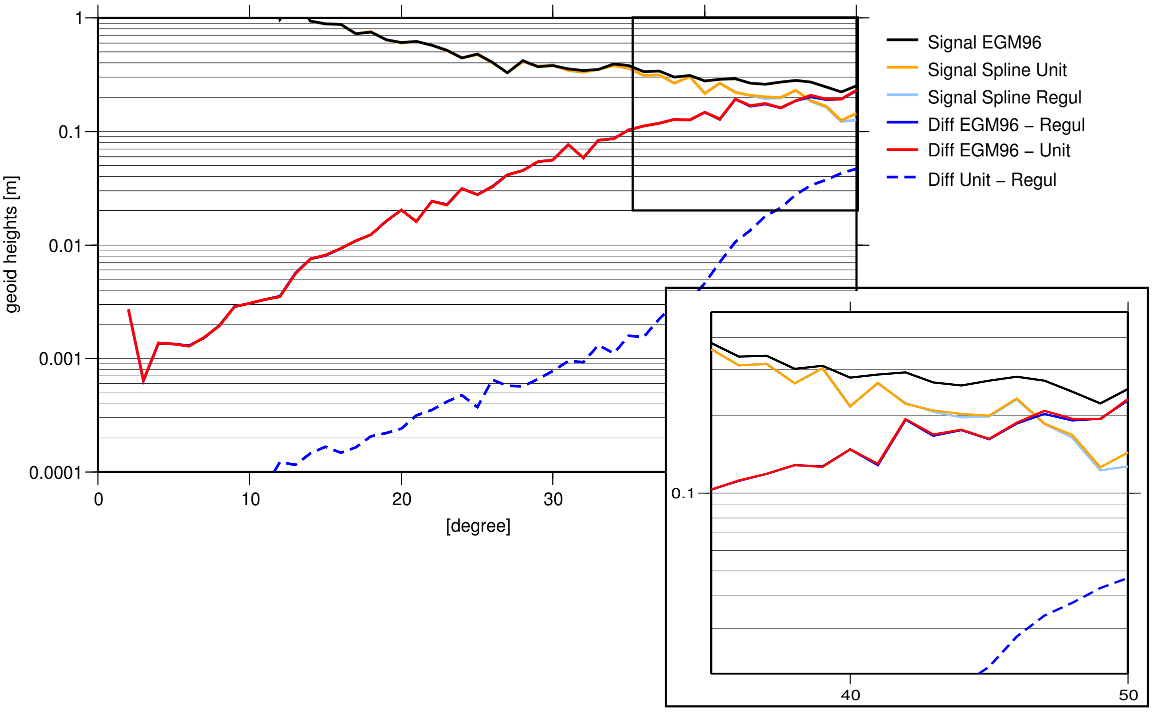


Figure 7.3: Degree variances of spline solutions, regularization by unit matrix compared to regularization with fully occupied regularization matrix, basis function unmodified, grid: triangle vertex level 16, $N = 50$

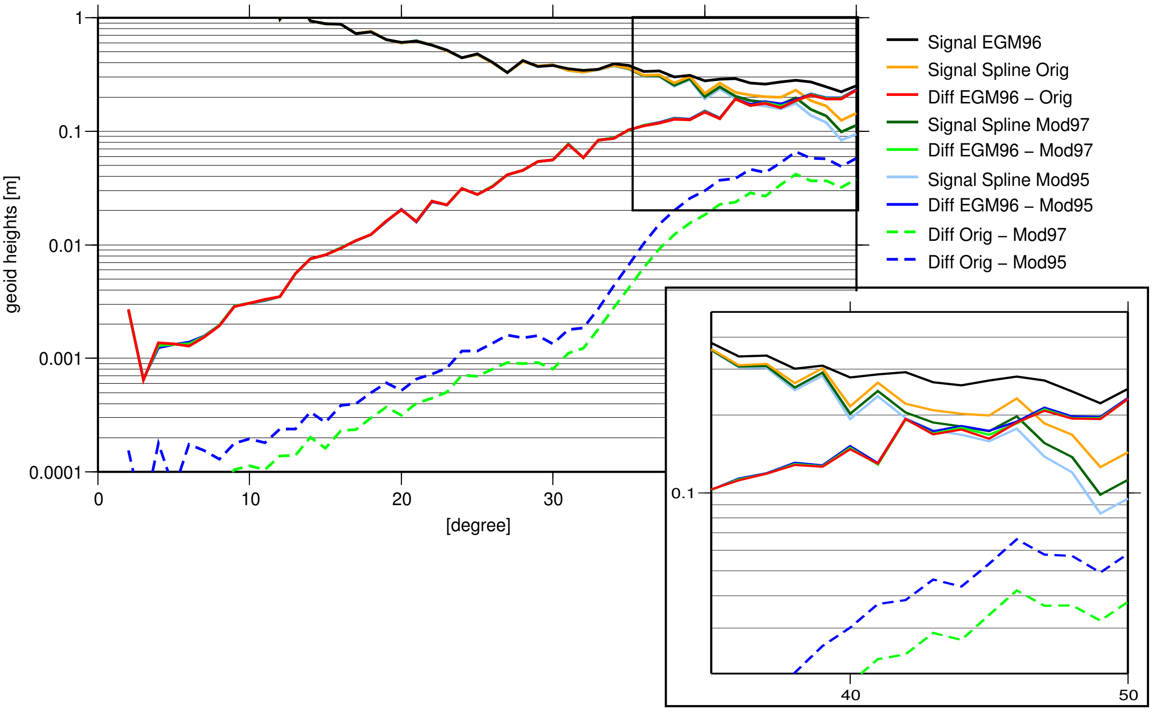


Figure 7.4: Degree variances of spline solutions, different modification factors applied to the spline kernels, regularization by unit matrix, grid: triangle vertex level 16, $N = 50$

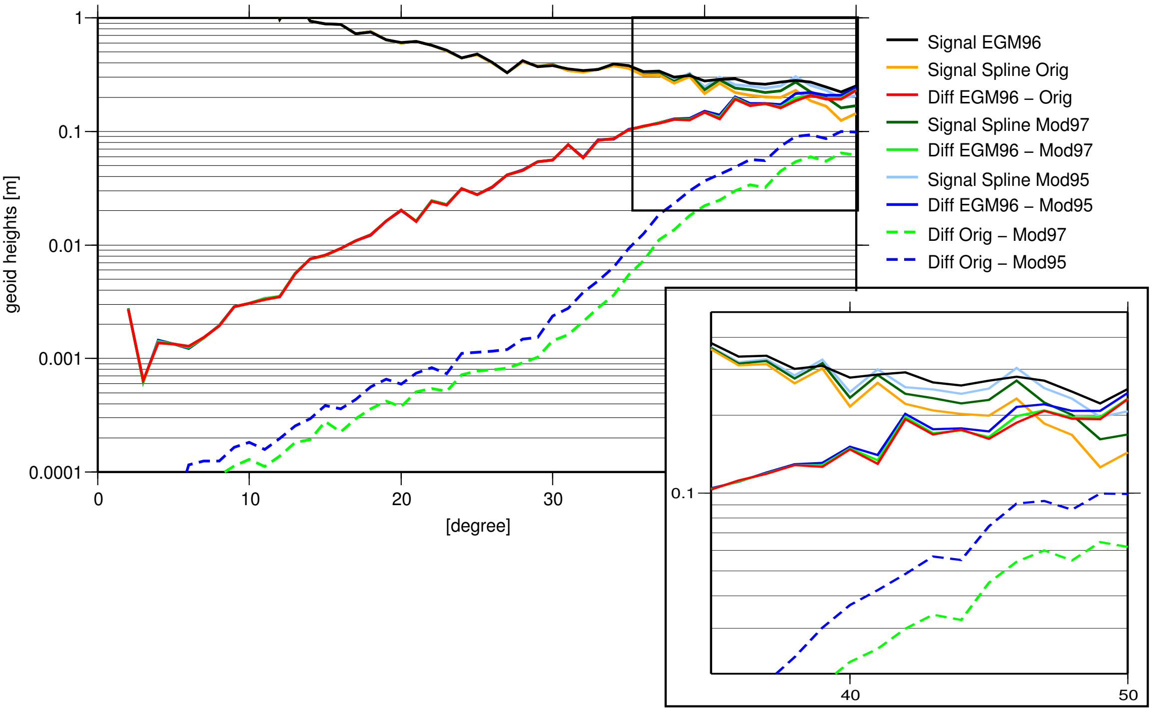


Figure 7.5: Degree variances of spline solutions, modified covariance functions, basis function unmodified, grid: triangle vertex level 16, $N = 50$

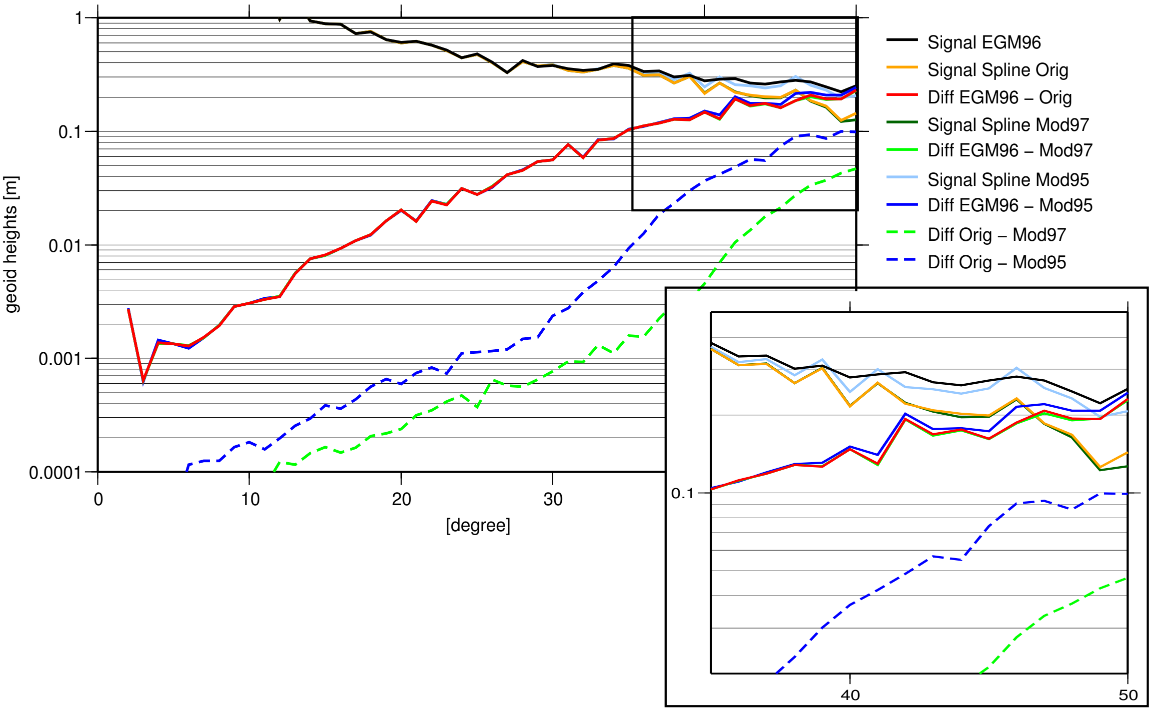


Figure 7.6: Degree variances of spline solutions, different modification factors applied, regularization by fully occupied regularization matrix, grid: triangle vertex level 16, $N = 50$

gravity field model, the regional refinements have to be calculated with global coverage, and the individual patches have to be combined to achieve a global solution. If desired, this global solution can then be converted to a parameterization in terms of spherical harmonics, this conversion being carried out by means of the exact quadrature methods, in this case the Gauss-Legendre quadrature, as described in Section 6.2.1. The different steps in the regional refinement procedure are summarized below:

1. Global reference field (generally modeled in terms of a spherical harmonic expansion)
2. Calculation of regional refinements (parameterized by space localizing basis functions)
3. Assembling of regional refinements with global coverage to obtain a globally refined solution (only necessary if a global gravity field solution is aspired)
4. Possibility of converting the globally refined solution to a spherical harmonic expansion by means of numerical quadrature methods

7.3.1 CHAMP

The regional gravity field recovery results from real CHAMP data will be presented in the following. The results are based on the article published by Eicker et al. (2004). The calculations are performed on the basis of the kinematical orbits derived by D. Švehla and M. Rothacher from the GPS observation of the CHAMP satellites, as presented in Švehla and Rothacher (2001) and Švehla and Rothacher (2003). The resulting data set comprises the data of one year, from 2002/03 to 2003/03. This data set was also used in the determination of the global CHAMP solutions ITG-Champ01s, ITG-Champ01k, and ITG-Champ01e, as presented by Mayer-Gürr et al. (2005) and described in more detail in Mayer-Gürr (2006). The various global solutions are distinguishable because of different choices of the regularization procedure, i.e. without any regularization, by applying Kaula's rule of thumb in the regularization matrix, and by introducing additional information provided by the EGM96, respectively.

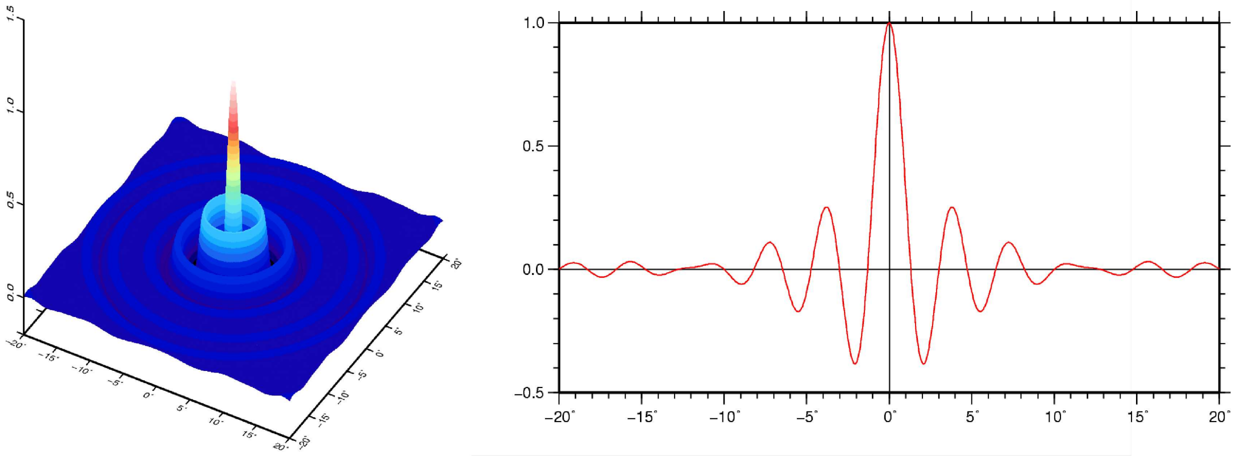


Figure 7.7: Spline kernel applied in the CHAMP gravity field recovery: developed until degree $n = 120$. As shape coefficients for the basis functions served the formal errors of ITG-Champ01k up to degree $n = 70$, above $n = 70$ Kaula's rule was used to approximate the coefficients.

The global solution ITG-Champ01K was chosen to serve as reference field in the regional analysis process. The coefficients of the global field were published up to degree $n = 70$ via the International Centre for Global Earth Models (ICGEM)[1], and this publicly available solution was introduced as reference field. The regional

---

[1] http://icgem.gfz-potsdam.de/ICGEM/ICGEM.html

gravity field is supposed to model the signal still present in the data after subtracting the reference field. As described in Section 3.4, the basis functions are constructed on the basis of the degree variances of the expected gravity field signal. Therefore, the formal errors of the global field ITG-Champ01k were applied up to degree $n = 70$, as they indicate the signal content that is not sufficiently modeled by the global solution. Above degree $n = 70$, Kaula's rule was used to approximate the frequency spectrum to be anticipated in the higher degrees, here up to degree $n = 120$. The resulting spline kernel is displayed in Fig 7.7. The regional refinements were supposed to be calculated with a resolution corresponding to a spherical harmonic degree $n = 120$ which features $121^2 = 14,641$ unknown parameters. To achieve a comparable resolution, the spline kernels were arranged according to a triangle vertex grid (as described in Section 3.5.1.6) of level 38, resulting in 15,212 global spline parameters. The mean nodal point distance defining the spatial resolution amounts to about 188 km. Concerning the question of the comparison of the resolution between a spherical harmonic expansion and a parameterization in terms of space localizing basis functions, refer to Section 3.5.3. Applying radial basis functions as parameterization, the regional solutions were calculated in the subsequent step. The regional areas were defined along a geographical grid, and the satellite data was restricted to the respective region, taking into account an additional strip of 5° width around the inner recovery region. The basis functions were arranged in an area slightly larger than the designated regional solution, here a border of 10° has proven to be reasonable. Further enlargement of either the satellite data area or the area containing the spline kernels did not provide significant improvements in the solutions. Thus it may be concluded that even though the regional gravity field may be influenced by mass anomalies of further distance to some extend, this influence seems to be possible to be neglected in the practical calculations.

The results for two exemplary regional refinement solutions are shown in Figure 7.8. Displayed is the additional information that was possible to be extracted from the data set. Obviously, especially in the area of the Andes and the Himalaya mountains, being regions of the Earth featuring very rough gravity field information, there is still signal present in the data. In Fig. 7.9 the regional refinement areas are chosen to provide a complete coverage of the surface of the Earth. The blocks have a size of $\Delta\lambda = 90°$ and $\Delta\vartheta = 70°$ with two spherical caps with an aperture angle of 40° covering the poles.

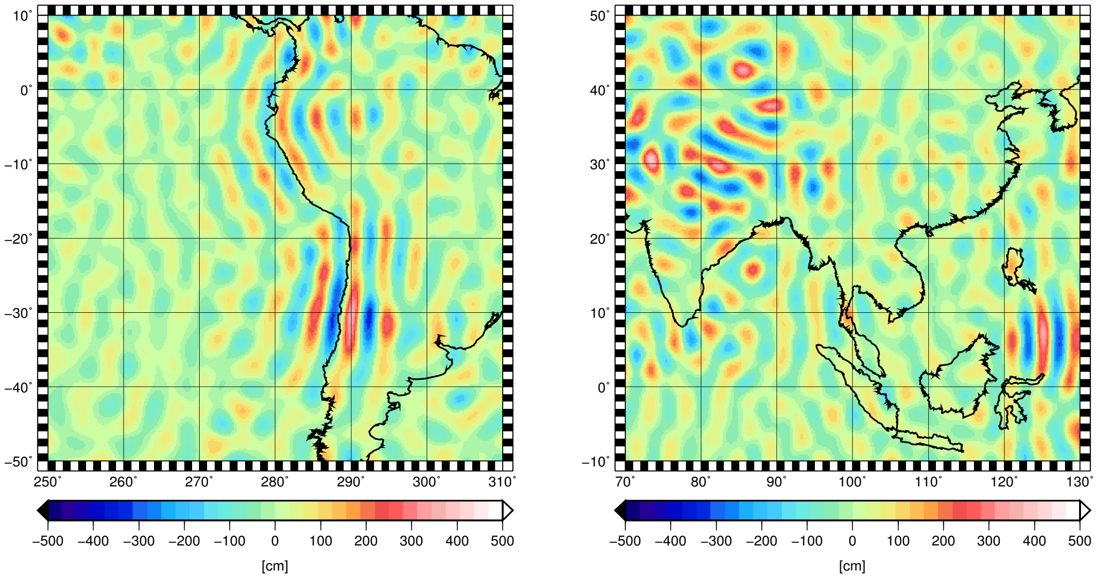


Figure 7.8: Regional refinements from CHAMP data: Andes and Himalaya region

Observing two adjacent patches reveals that they match very well despite the individual recovery of each regional solution. By means of the Gauss-Legendre quadrature, the refinement patches can be merged to obtain a global spherical harmonic expansion. In order to apply the quadrature formula, the gravity field functionals have to be evaluated at the required quadrature nodes. This does not, however, pose any difficulties, because from the continuous spline representation the modeled functional can be evaluated at arbitrary points on the sphere. The validation of the refined solution was performed by comparison with a GRACE gravity solution, in this case the GGM01s (Tapley et al. 2004), which was calculated from 111 days of GRACE data. The results obtained from GRACE data can be presumed to provide superior accuracy, and therefore the quality of a CHAMP gravity field can be measured by its differences to the GRACE solution. In order to perform the comparison, first the differences between the original reference field ITG-Champ01k and the GGM01s were calculated and are shown in Fig. 7.10. The RMS of the differences amounts to 96.57 cm with a maximum deviation of 1031.04 cm. Subsequently, the regional refinement was added to the reference solution, and again the differences compared to the GGM were determined, as displayed in Fig. 7.11. The RMS improved to a value of 85.54 cm, and the maximum difference was reduced to 920.30 cm. The impact of the regional refinement can, therefore, be specified by a significant improvement in accuracy of about 10%.

Because of the conversion from the spline solution to a spherical harmonic expansion, the results can be investigated in the frequency domain as well. Even though the field ITG-Champ01k was published only up to degree $n = 70$, it was originally calculated up to degree $n = 90$. Internal access to this solution allows the examination up to the higher resolution. Fig. 7.12 shows the degree variances in terms of geoid heights for the difference between the ITG-Champ01k and the GGM01s (blue) and for the difference between the refined solution and the GGM01s (red). The directly calculated global solution exhibits smaller differences in the lower frequency part, approximately up to degree $n = 50$. The higher accuracy in these long wavelengths would require an extra careful handling of the transitions between refinement patches. In the higher frequency part, however, the regionally refined solution performs better than the original solution, as can be observed from the red curve running slightly below the blue curve above degree $n = 50$. The differences of the original solution intersect with the signal curve (black) at degree $n = 90$, whereas the refined solution seems to contain information even in the frequency part above degree $n = 90$. From these investigations, it can be concluded that restricting the original spherical harmonic solution to degree $n = 70$ seems to have been too pessimistic, since Fig. 7.12 reveals the possibility of extracting additional information from the global solution above this degree. This discrepancy partly explains the large improvements in Fig. 7.11 when compared to Fig. 7.10. Nevertheless, the solution calculated from regional refinements still performs better in the higher frequencies than the global solution expanded up to degree $n = 90$, as can be observed from the degree variances plot. Also in the space domain, the global solution up to degree $n = 90$ provides an RMS of 87.26 cm and a maximum error of 1021.68 cm when compared to the GGM01s, which again are larger than the differences calculated for the regionally refined solution.

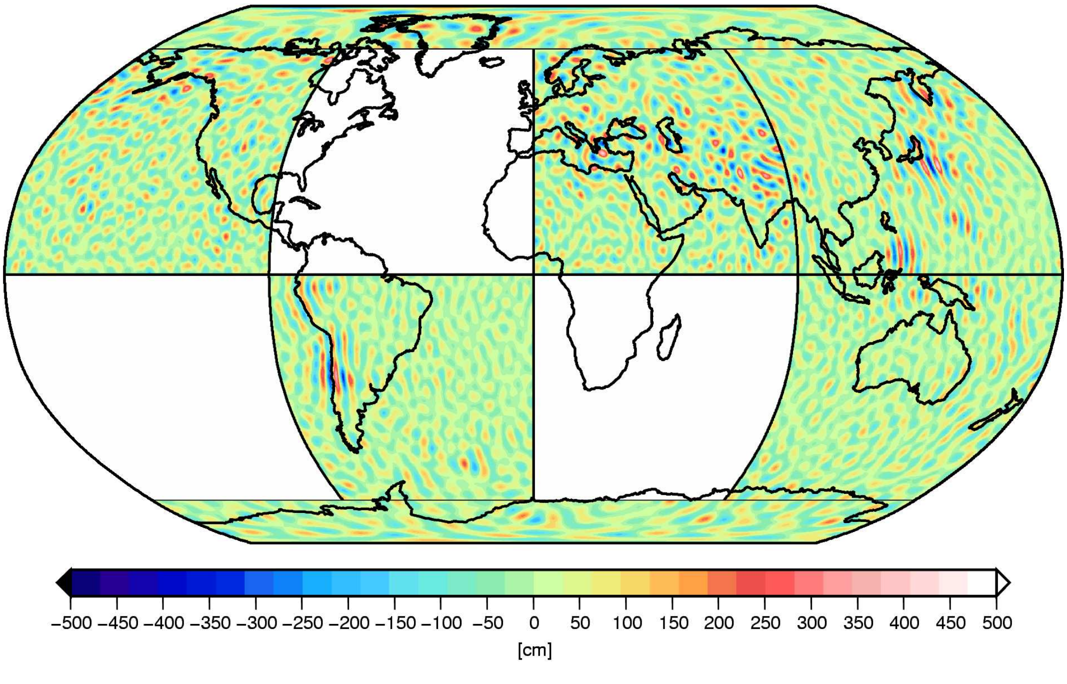


Figure 7.9: Individual refinement patches providing a global coverage of the Earth’s surface

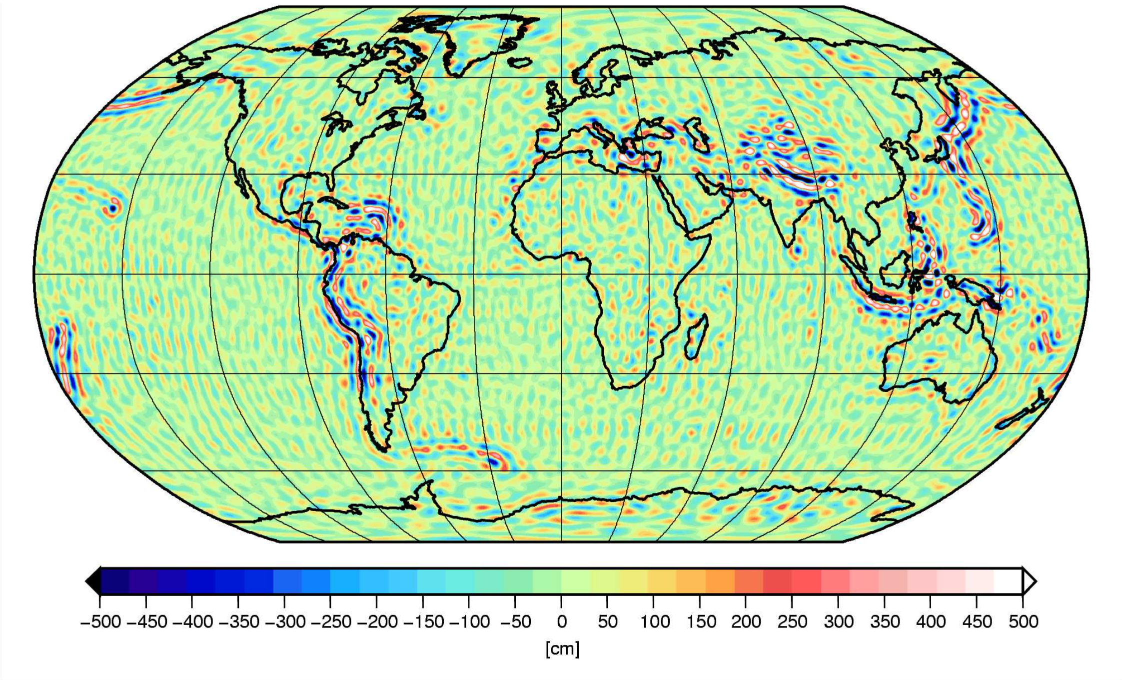


Figure 7.10: Differences between ITG-Champ01k ($N = 70$) and GGM01s (compared up to degree $N = 120$), RMS: 95.57 cm, Max: 1031.04 cm

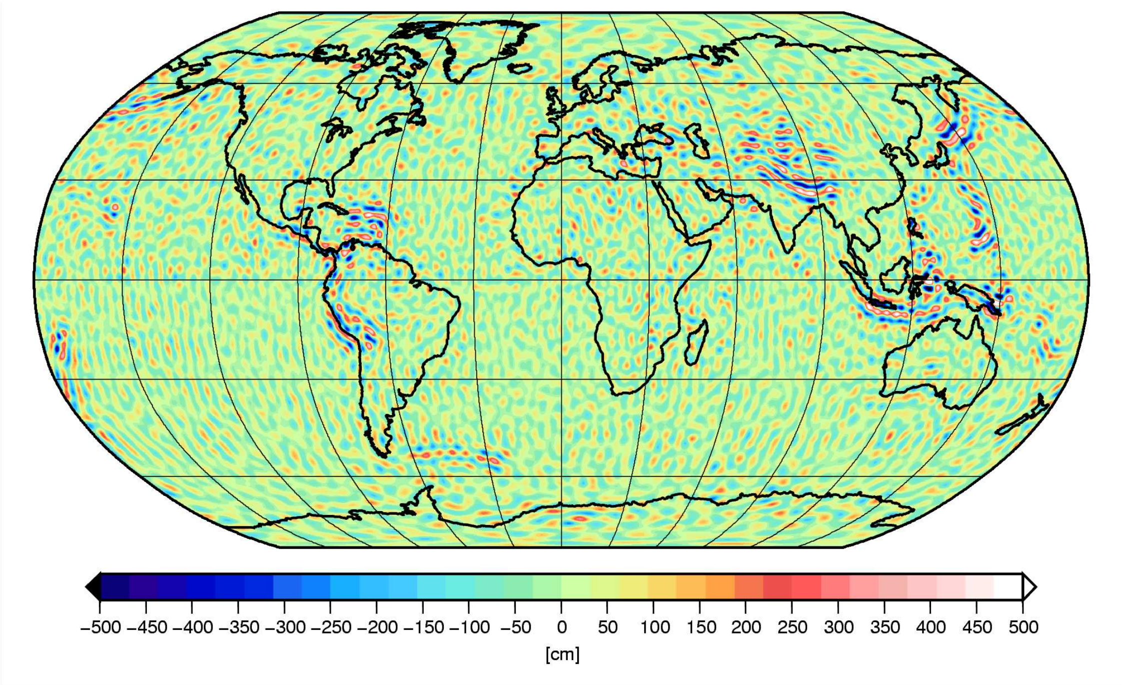


Figure 7.11: Differences between ITG-Champ01k ($N = 70$) refined by regional focus and GGM01S (compared up to degree $N = 120$), RMS: 85.54 cm, Max: 920.30 cm

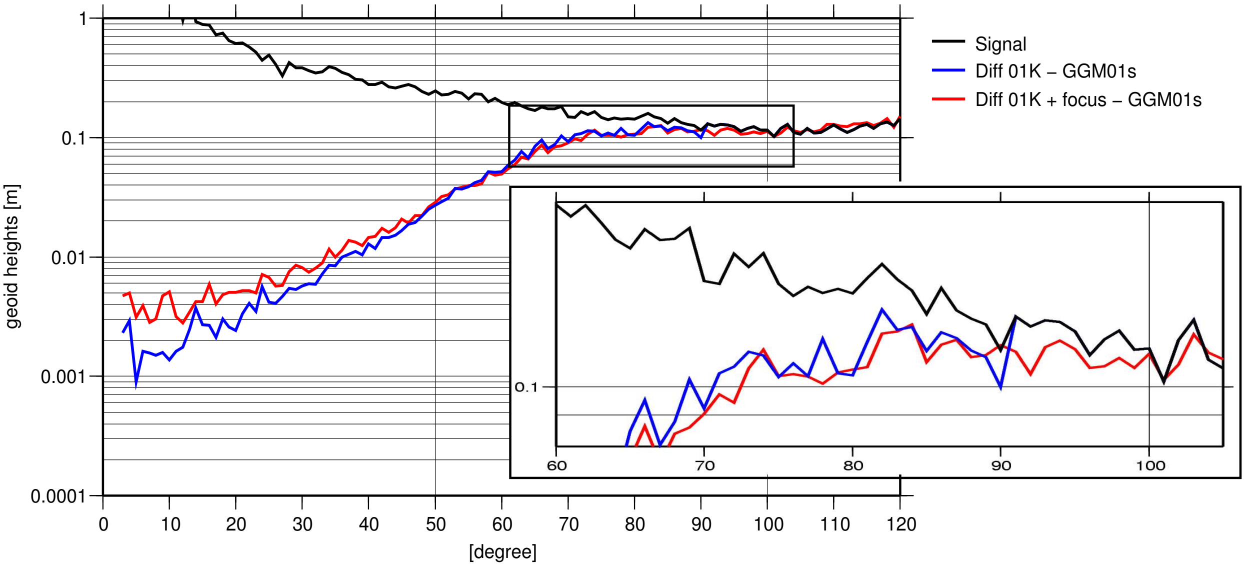


Figure 7.12: Differences in degree variances (in terms of geoid heights) between ITG-Champ01k and GGM01s (blue) and between the refined solution and the GGM01s (red)

7.3.2 GRACE

7.3.2.1 One-month Solution

The satellite mission GRACE provides gravity field solutions with an accuracy that has never before been achieved for a global model. Therefore, the validation of the results represents a certain challenge, as no existing gravity field solution can serve as 'true' pseudo-real comparison. A way out of this problem can be the calculation of solutions from only a limited amount of data, i.e. from only a short period of time, and the use of solutions calculated from a larger amount of data as references. This procedure allows to make qualitative valuations about the applied approach. This is why, in the following scenario, a regional solution was calculated from only one month of GRACE data, to be more specific, from the data collected in June 2005. The results were then compared to the global GRACE gravity model ITG-Grace02s (Mayer-Gürr et al. 2007a). It was derived from three years of GRACE data, therefore it is expected to be of superior accuracy and can thus serve as comparison. As investigation area was chosen a region containing a large part of South America and part of the adjoining Pacific Ocean, the area being limited by the geographical coordinates $\lambda = [250°, 310°]$ and $\varphi = [-50°, 10°]$. It was selected because of its very pronounced high frequency signal in the Andes region and a comparably rather smooth signal in the respective part of the Pacific. These are conditions that allow the exploration of the strengths provided by the regional recovery approach particularly well. In geographical areas with strongly varying spectral behavior, the concept of the regionally adapted regularization procedure described in Section 5.2.5 can be especially helpful. In the proposed evaluation area, the partitioning into one regularization group for the continental area and one for the ocean area, as given in Eq. (5.51), seems to be a reasonable choice. As reference field served the EGM96 (Lemoine et al. 1998) up to degree $n = 240$, and the regional solutions were calculated as refinements to the reference field. For the radial basis functions to model this residual signal content, the error degree variances of the reference field were used as shape coefficients in the series expansion of the spline kernels, which was performed up to $n = 140$. The basis functions were located at a triangle vertex grid of level 45, which results in a resolution comparable to a spherical harmonic degree $n = 140$.

The gravity field recovery was performed using the integral equation approach, as specified in Section 4.3.2. The results obtained from the setting described above are presented in Fig. 7.13. Illustrated are the differences between the regionally refined solution compared to the ITG-Grace02s up to degree $N = 140$. The left hand side of this figure shows the outcome in the case of one uniform regularization parameter determined for the region with an RMS of the differences amounting to 16.51 cm. This is compared to the differences for the adapted regularization procedure illustrated on the right hand side of Fig. 7.13. The introduction of the additional regularization area leads to an improvement of the RMS to 12.86 cm, which is significantly smaller. To evaluate the quality of the one-month regional spline solution, it was tested against the same month of the GFZ-RL04 (Flechtner 2007) monthly gravity fields. In a first step, both monthly solutions were again compared to the ITG-Grace02s. In this case the comparison was only performed up to degree $N = 120$, as the GFZ-RL04 series is published only up to this resolution. The results are presented in Fig. 7.14, for the spline solution on the left side and for the GFZ-RL04 model on the right side. The relation of the RMS values of 6.6 cm and 13.9 cm, respectively, reveal that the spline solution matches the ITG-Grace02s model better than is the case for the GFZ-RL04 solution. The examination of the associated figures also exhibits the error pattern to be less pronounced in case of the spline solution. Since the ITG-Grace02s and the regional refinement model were calculated using the same analysis procedure, this might not be a fair comparison, however. Therefore, the same differences were calculated between the two monthly solutions and the Eigen-GL04C (GFZ Potsdam 2007). This global model was derived from 2.5 years of GRACE observations supplemented by additional surface gravity data. The results are compared in Fig. 7.15; again it becomes obvious that the spline solution shows better agreement with the global model than is the case for the GFZ-RL04 solution. The RMS values of 9.3 cm in contrast to 16.5 cm also support this conclusion.

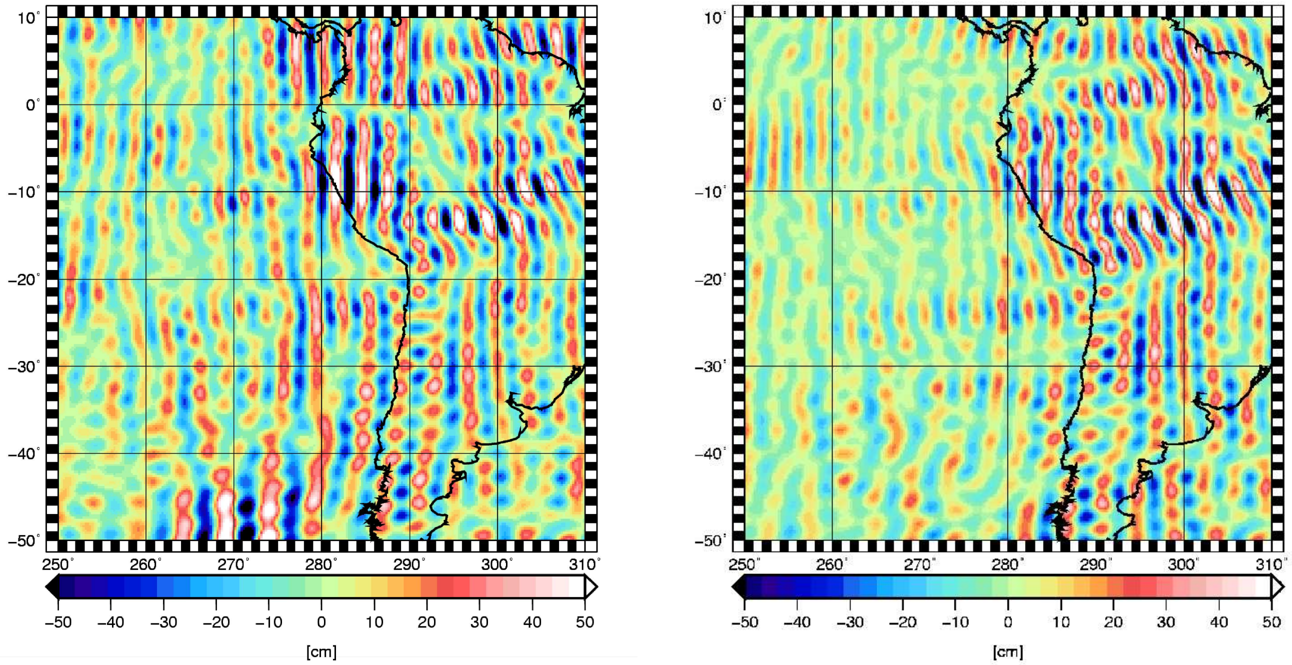


Figure 7.13: Differences between spline solution (05-2006) and ITG-Grace02s, one uniform regularization parameter (left, RMS: 16.51 cm) and adapted regularization (right, RMS: 12.86 cm), $N = 140$

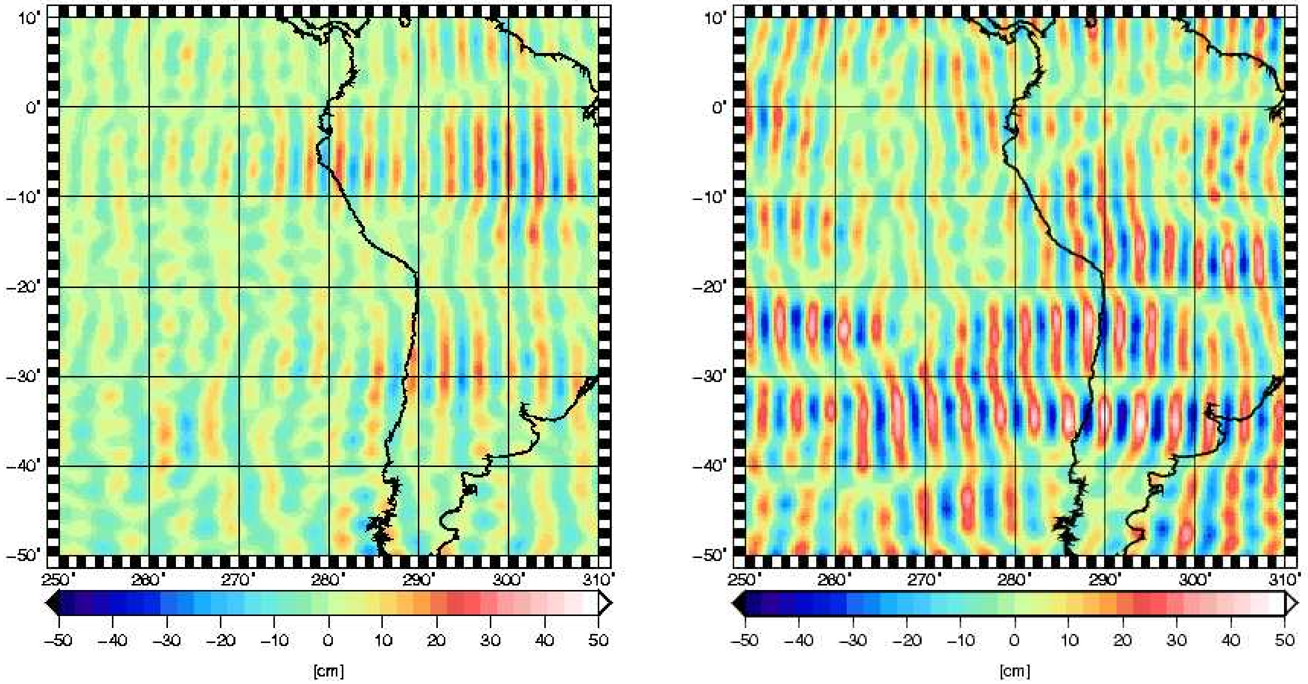


Figure 7.14: Differences between spline solution (05-2006) and ITG-Grace02s (left, RMS: 6.6 cm) and differences between GFZ-RL04 (05-2006) and ITG-Grace02s (right, RMS: 13.9 cm), $N = 120$

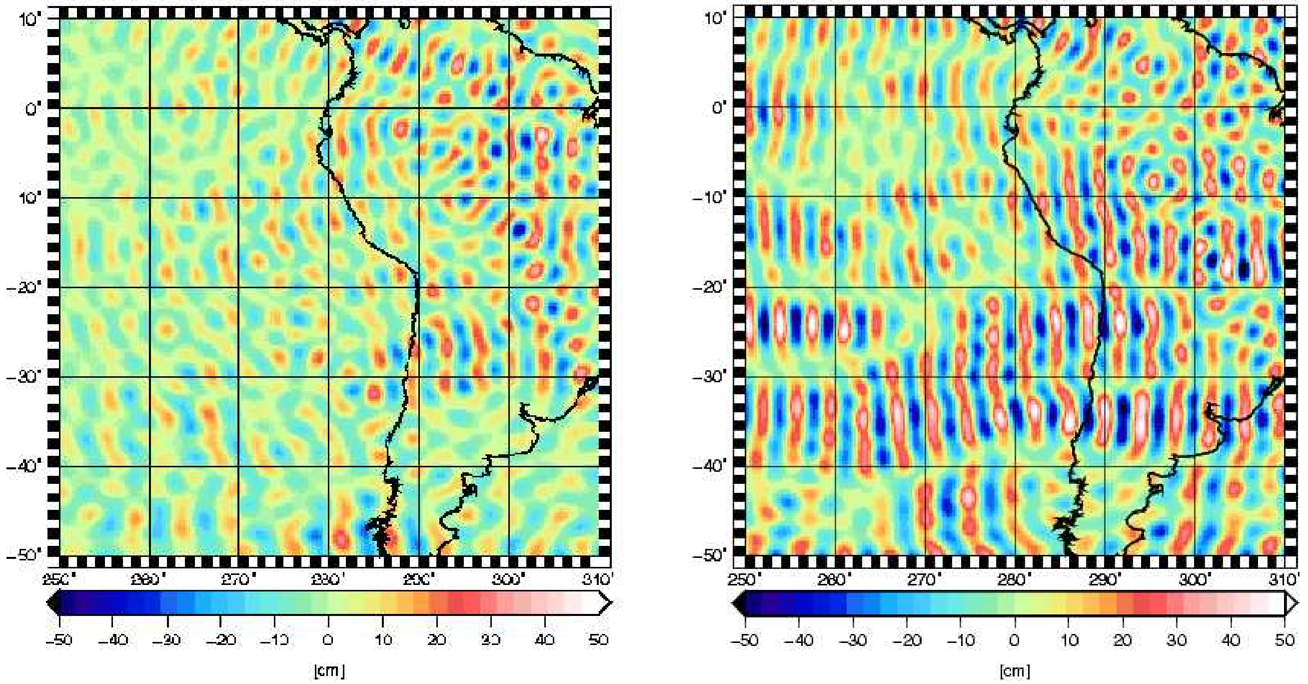


Figure 7.15: Differences between spline solution (05-2006) and Eigen-GL04C (left, RMS: 9.3 cm) and differences between GFZ-RL04 (05-2006) and Eigen-GL04C (right, RMS: 16.5 cm), $N = 120$

7.3.2.2 ITG-GraceSpline03

This section is dedicated to the regional refinement of the global gravity field ITG-Grace03s (Mayer-Gürr et al. 2007b), the latest GRACE model calculated at the IGG Bonn. The model comprises a static gravity field, represented by a spherical harmonic expansion up to degree $n = 180$ and a time variable part with a resolution of up to $n = 40$. In the time domain it is parameterized either by monthly means or by quadratic splines. In the following, only the static model will be taken into consideration. It was computed from accumulated normal equations over a time span from September 2002 to April 2007, using GRACE data only and without applying any a-priori information or regularization. Concerning the functional model, the integral equation approach, as described in Section 4.3.2, was applied. Using ITG-Grace03s as a global reference model, the regional refinement strategy was supposed to extract additional gravity field information from the same data set using exactly the same analysis procedure as in the case of the global solution. The reference field was restricted to a spherical harmonic degree of $n = 145$, as the higher degrees are too severely corrupted by noise. For the construction of the radial basis function, the formal errors of ITG-Grace03s were used up to $n = 145$, as they are assumed to reflect the amount of information for each degree that has not yet been modeled by the global solution. Above, the series expansion was approximated by Kaula's rule up to $n = 180$. The respective basis function is plotted in Fig. 7.16. The spline kernels were located at the nodes of a grid resembling the resolution of $n = 180$, that is a triangle vertex grid of level 57 with 33,642 nodal points. Resulting from this specification, the spatial resolution of the spline solution averages about 123 km nodal point distance. The surface of the Earth was divided into patches of size $\Delta\lambda = 90^\circ \times \Delta\vartheta = 60^\circ$ with a spherical cap featuring an aperture angle of 60° at each of the poles. For each of the areas, a regional refinement was calculated, again using the regionally adapted regularization procedure as introduced in Section 5.2.5 with individual regularization parameters for continental and oceanic regions. From the individual spline solutions, a global spherical harmonic solution was derived by means of the Gauss-Legendre quadrature. As explained in Section 6.2.1.2, a number of $L = 181$ circles of latitude is sufficient to obtain spherical harmonic coefficients of up to degree $n = 180$.

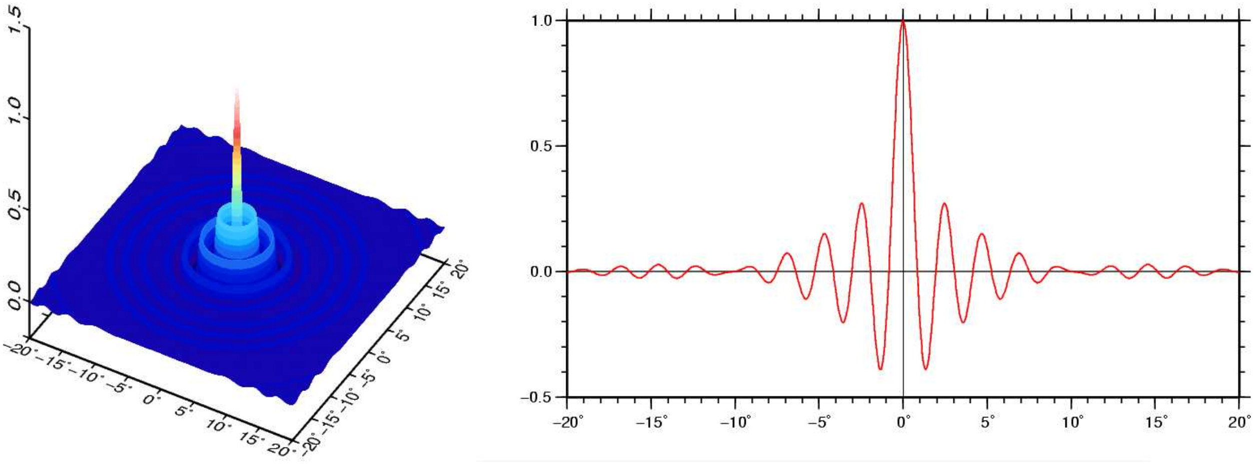


Figure 7.16: Spline kernel applied in the refinement of the global model ITG-Grace03s: developed until degree $n = 180$, formal errors of ITG-Grace03s up to degree $n = 145$, above $n = 145$ Kaula's rule

This global solution calculated from individually determined refinement patches is denoted by ITG-GraceSpline03, and it is illustrated and evaluated in Fig. 7.17 to Fig. 7.19. Concerning the evaluation or validation of the obtained solution, there arises the question of a proper comparison. GRACE has delivered gravity field solutions with an accuracy in the long and medium wavelength part of the spectrum that has never been achieved before. This makes validation a challenging task. The regional refinement process aims particularly at the improvement of the higher frequencies, therefore terrestrial gravity field information can be used for evaluation purposes. In the results described here, the global model Eigen-GL04C (GFZ Potsdam

2007) was chosen as comparison, as the terrestrial gravity data taken into account in addition to the GRACE data allows the Eigen-GL04C to be assumed of exhibiting superior accuracy in the higher frequency part of the gravity field spectrum. The comparison in the frequency domain is shown in Fig. 7.17 by displaying degree variances in terms of geoid heights. The signal of Eigen-GL04C is thereby visualized by the black curve. The ITG-Grace03s and ITG-GraceSpline03 are represented by the blue and red curves, respectively, the dashed lines stand for the signal and the continuous lines illustrate the differences compared to the Eigen-GL04C. The difference between the two ITG solutions is indicated by the green curve. Examining the error curves reveals that the two directly calculated spherical harmonic solutions Eigen-GL04C and ITG-Grace03s match better in the lower frequency part, up to about degree $n = 70$. The extremely high accuracies in the long wavelengths would require an additionally careful treatment of the transitions between adjacent refinement patches. As expected, the benefit of the regional refinements becomes evident especially in the higher degrees, in the case at hand starting from approximately $n = 148$. From this point on, the blue error curve runs significantly above the red one. The blue line intersects the black line at about $n = 165$, which implies that above this degree the errors are larger than the signal itself. Therefore, above $n = 165$ no further gain of information can be expected from the directly calculated global solution ITG-Grace03s. Since this model was calculated without any regularization, the amplified errors dominate the respective signal curve (dashed blue line) in the high frequencies, leading to an unrealistically large signal above $n = 140$. The noticeable jump in the error curves (in the red as well as in the blue one) around degree $n = 115$ originate from an inadequate weighting of the satellite data and the terrestrial gravity information in the calculation of the Eigen-GL04C. Regarding ITG-GraceSpline03, there seems to be full signal present in the solution up to almost degree $n = 160$. In the higher degrees the dashed red signal curve becomes significantly dampened due to the regularization process. But the intersection of the red error curve and the black signal curve does not occur before approximately $n = 180$, thus up to this resolution there seems to be information remaining in the regionally refined solution. The obtained results can be evaluated in the space domain as well. Therefore, the differences in geoid heights were calculated between the two ITG solutions and the Eigen-GL04C, the respective plots are displayed in Fig. 7.18 for ITG-Grace03s and in Fig. 7.19 for ITG-GraceSpline03. As there

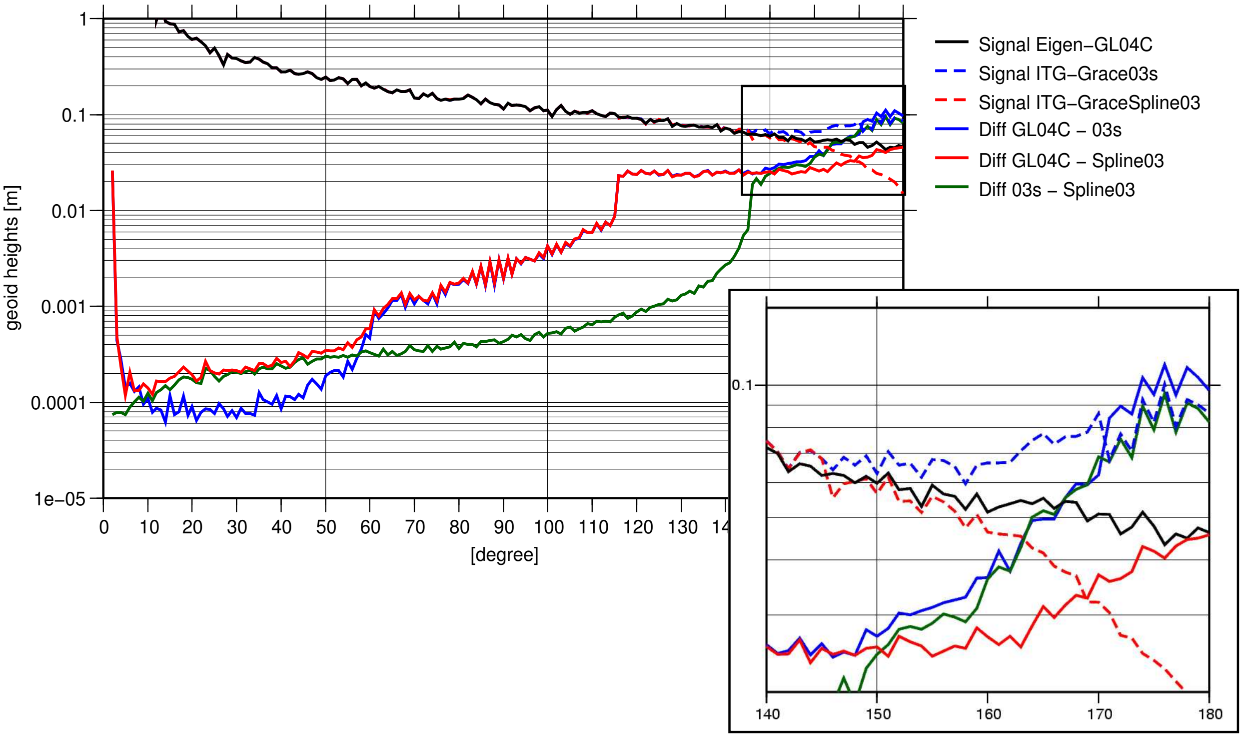


Figure 7.17: Degree variance of ITG-Grace03s and ITG-GraceSpline03 and the differences of both fields compared to the Eigen-GL04C

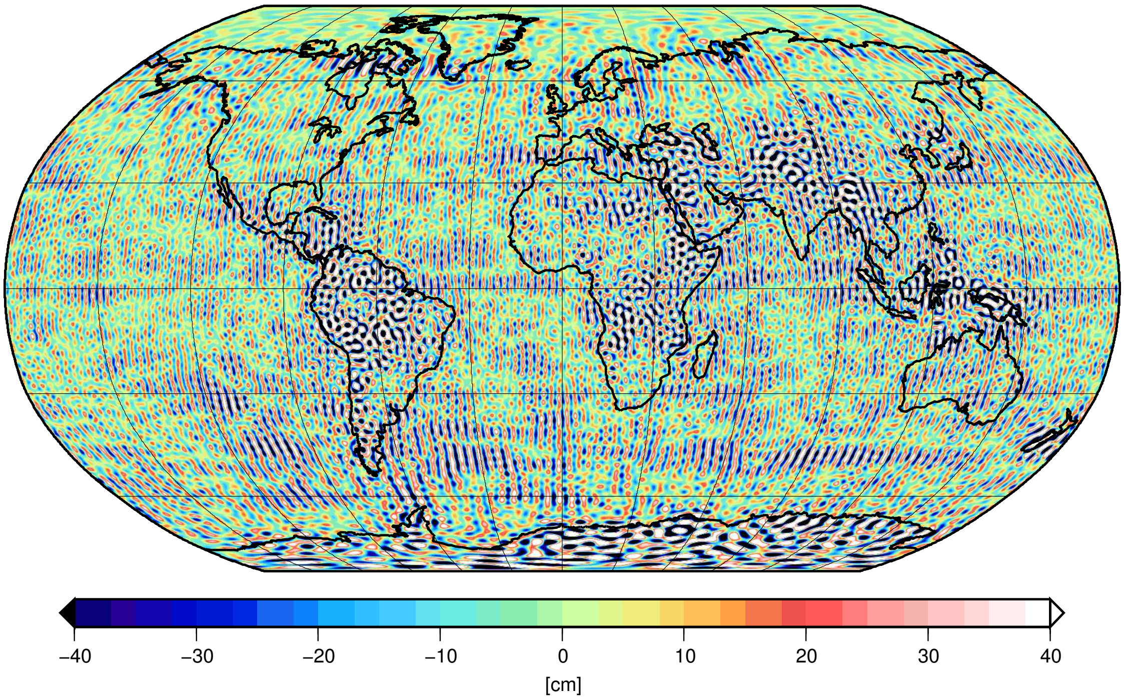


Figure 7.18: Differences between ITG-Grace03s and Eigen-GL04C, $N = 165$, RMS: 20.6 cm

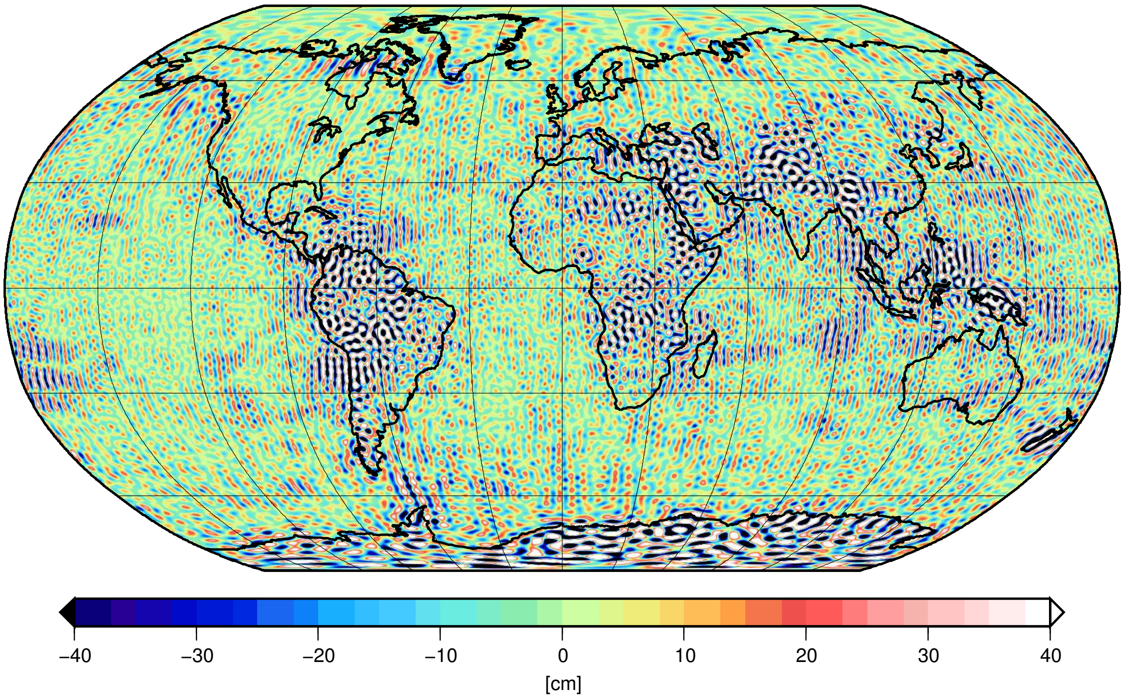


Figure 7.19: Differences between ITG-GraceSpline03 and Eigen-GL04C, $N = 165$, 17.9 cm

seems no signal to be provided above $n = 165$ by the directly determined global solution, the differences were only calculated up to this resolution. A comparison of both figures reveals that the regional refinements have indeed provided significant improvement. The RMS was reduced by about 13% from 20.6 cm to 17.9 cm supporting this conclusion. It has to be mentioned, however, that the model ITG-Grace03s was calculated without any regularization applied, resulting in an amplification of the errors in the high frequencies. Possibly, a regularization process would provide enhancement in the resolution of the global model as well. But nevertheless this would only imply a uniform global dampening of the gravity field features. Therefore, it can be assumed that, in comparison to a globally regularized model, the regionally refined solution would provide a further gain in accuracy and resolution, too.

The results provided by GRACE commonly exhibit a typical error pattern of distinct stripes in north-south direction. This is evident in Fig. 7.13 to 7.15 as well as in Fig. 7.18 and 7.19. Therefore, concluding remarks are to be made concerning the nature and causes of this error structure. The following reasons can be speculated:

- The primary observation, the K-band range or range-rate measurements, are carried out in along-track direction, thus the constraints given by the measurements are strong in this direction. As the measurement errors are always directed into the direction in which the model offers the largest degree of freedom, they are directed to across-track direction causing the stripe pattern.
- The second effect leading to a stripe pattern in the error display is the fact that the background models (such as ocean tide model, atmosphere model, etc.) are of insufficient accuracy compared to the superior accuracy provided by the GRACE K-band instrument. The unmodeled short periodic gravity field changes cause aliasing effects. These are characterized by meridional stripes, because the observations are sampled along the orbit by a satellite flying in north-south direction. This results in arcs that are close to each other in space do not necessarily have to be close in time as well. Between neighboring arcs there may have passed a considerable period of time due to the orbit configuration of the GRACE mission; during this time the situation in the (unmodeled) background models may have changed completely. Investigations concerning these aliasing effects have been carried out, for example, by Seo et al. (2008).

7.3.3 Combination of GRACE and GOCE

Due to the lack of real GOCE observations, the following calculations were performed in the frame of a simulation scenario. The simulation example is supposed to demonstrate that a regional refinement strategy cannot only be applied to refine global solutions, but it can be utilized to combine different data sets as well. In the scenario at hand, observations of GRACE are supposed to be combined with those of GOCE. The GRACE solution serves as the global reference field, as GRACE provides excellent results especially in the long and medium wavelength part of the gravity field spectrum. The regional refinements of this solution are then calculated on the basis of GOCE-like observations, as they cover the short periodic part of the gravity field spectrum with superior accuracy. As additional reference field for the higher frequencies, the OSU91 (Rapp et al. 1991) was introduced from degree $n = 150$ to $n = 300$. The procedure was described in Eicker et al. (2005) and Eicker et al. (2006). An extension of this combination approach is suitable to include terrestrial or airborne gravity data as well.

In order to achieve a consistent data set, both the GRACE and the GOCE solution were calculated from simulated observations on the basis of the EGM96 (Lemoine et al. 1998) up to degree $n = 300$. The observations for GRACE as well as for GOCE were simulated for a period of 30 days with a sampling rate of 5 sec. In case of the GRACE satellites, the simulated orbit positions were corrupted by a white noise with a standard deviation of 3 cm and the intersatellite ranges between the GRACE twin satellites with a white noise of 10 μm. Regarding the GOCE satellite, the gradiometer observations were corrupted by a colored noise model with a standard deviation of 1.2 mE. In this model the PSD is assumed as constant in the measurement band of the gradiometer and features increasing energy in the long wavelength part of the spectrum, as the gradiometer is not able to recover these low frequencies. For the satellite positions again an accuracy of 3 cm was assumed. From this simulated data set, first the GRACE solution up to a spherical harmonic degree of $n = 150$ was derived. The analysis of the GRACE data was conducted using the integral equation approach described in Section 4.3.2. The regional refinements from GOCE observations were then calculated as residual fields to this global solution, applying the functional model indicated in Section 4.3.3. Using a parameterization by splines as space localizing base functions, the spline kernels are to represent the gravity field features to be determined. Therefore, up to degree $n = 150$, the error degree variances of the (simulated) GRACE solution were applied as coefficients in the series expansion of the spline kernel. These error degree variances represent the signal which is still in the data in addition to the GRACE solution. Above degree $n = 150$, the degree variances were approximated by Kaula's rule of thumb. The resulting basis function is displayed in Fig. 7.20. The basis functions were located at the nodes of a triangle vertex grid of level 95. This results in 92,162 global points, guaranteeing a resolution slightly higher than a spherical harmonic degree of $n = 300$. The resulting grid pattern exhibits an average nodal point distance of about 74 km. From the setting described above regional solutions were calculated. The results for an individual regional refinement are displayed in Fig. 7.21 for the area of the Andes. The figure shows the differences to the pseudo-real field EGM96 in terms of geoid heights. The comparison was only performed up to degree $n = 240$, because the higher degrees are too strongly corrupted by noise. Here again the impact of the regionally refined regularization procedure becomes evident, as the RMS value improves from 9.24 cm (left side, one uniform regularization parameter) to 8.08 cm (right side, adapted regularization for ocean and continental area). This corresponds to an improvement of 12.5%. Subsequently, regional refinements were calculated covering the complete surface of the Earth, with an overlapping border of 10° having been taken into account at the boundaries to avoid truncation effects. The individual patches have a size of $\Delta\lambda = 40° \times \Delta\varphi = 50°$ complemented by two spherical caps with an aperture angle of 30° covering the poles. The pattern of the refinement regions is illustrated in Fig. 7.22. It leads to a number of about 5000 to 9000 spline parameters to be determined for each region, the size of the patches being limited by storage restrictions. From the individual spline solutions the gravity field functionals were predicted to the nodes of a Gauss-grid with a spacing of $\Delta\lambda = 0.5°$. This corresponds to a number of 360 circles of latitude. From the global field, a spherical harmonics expansion was calculated using the Gauss – Legendre quadrature. The results are shown in Fig. 7.23, again compared up to degree $n = 240$. The global RMS amounts to 6.51 cm, including the poles. Here only the solution with the adapted land/ocean regularization is presented, as on a global scale the improvements are not as easily detectable, especially as for many of the patches the choice of continental and oceanic regions does not pose an ideal option. Here a more tailored adjustment of the regularization areas would be advisable. The global solution can be evaluated in the frequency domain as well; the corresponding degree variances are displayed in Fig. 7.24. The blue line specifies the difference degree variances between the global reference field and the EGM96, the dark green and red line indicate the differences of the refined

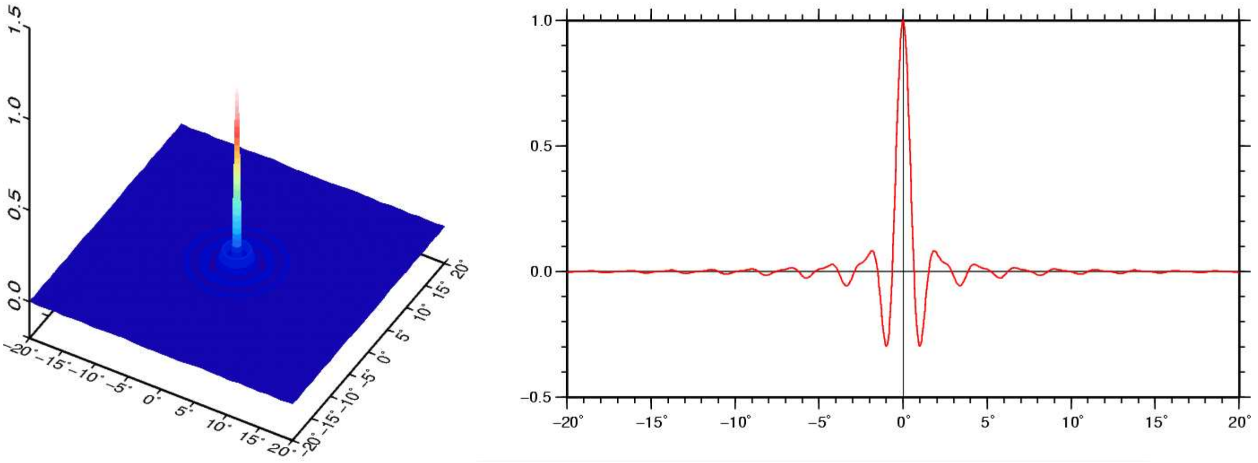


Figure 7.20: Spline kernel applied in the combined GRACE-GOCE simulation scenario: developed until degree $n = 300$, formal errors of simulated spline solution up to degree $n = 150$, above $n = 150$ Kaula's rule

solution using one or two regularization parameters, respectively. Naturally, the refined curve runs below the reference curve due to the additional information provided by the GOCE observations. Taking a closer look at the high frequency part reveals that even though the choice of the regularization areas is not perfectly tailored, the errors of the solution with adapted regularization are constantly smaller.

It shall be pointed out that the direct computation of the spherical harmonic coefficients by solving the improperly posed downward continuation cannot provide a stable solution up to an (arbitrarily) high degree as it can be achieved by means of the quadrature method. In this application, the maximal degree should be limited only by the signal content of the gravity gradient measurements. Furthermore, the regional recovery procedure offers a chance to deal with the polar gap problem in a tailored way, as in regions without any data the regional refinement can either be skipped or the regularization parameter can be adjusted accordingly.

The concept of taking a global solution and calculating regional refinements to the global solution presents an elegant way of combining different data sets. The lower resolution data set can serve as global reference, and observations with assumed higher resolution can be exploited for the regional refinements. The regularization parameters are determined individually for different geographical areas by variance component estimation, as explained in Section 5.2.4. This procedure optimally chooses the regularization parameter according to the given signal-to-noise ratio in the respective regularization area. In this way the dampening can be adjusted in accordance with the residual signal content (i.e. the information additional to the global reference solution).

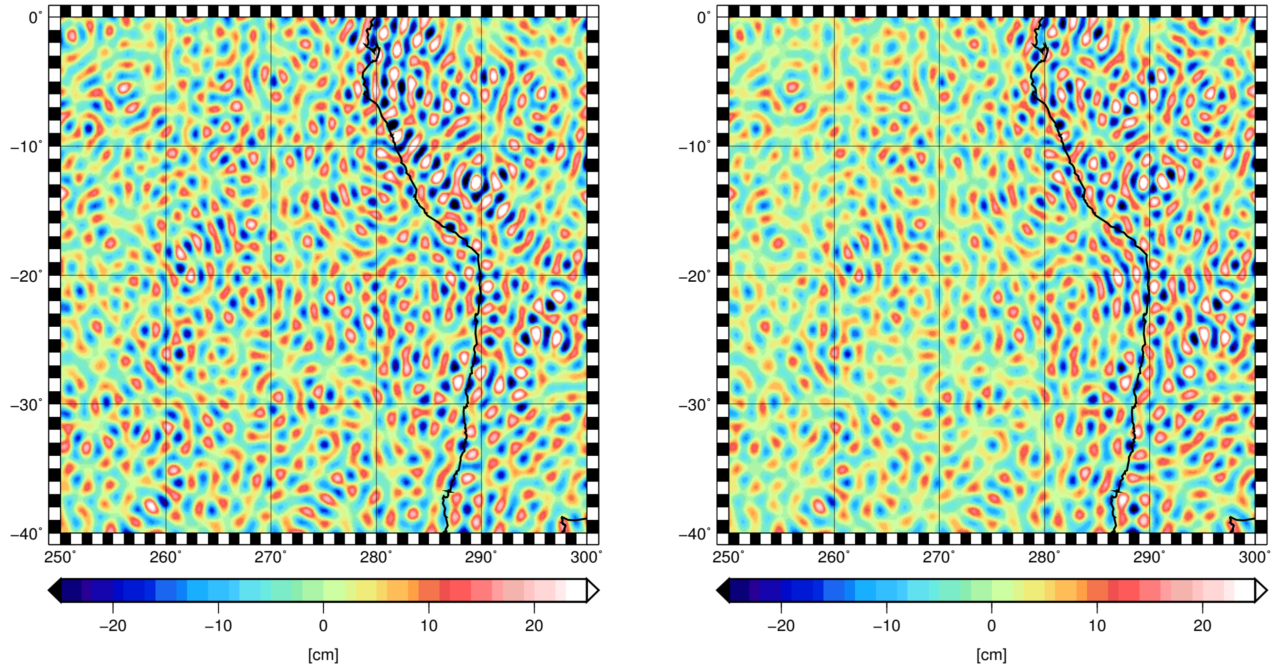


Figure 7.21: Differences between spline solution from combined GRACE/GOCE analysis (simulation study) and the pseudo-real field EGM96: one uniform regularization (left, RMS: 9.24 cm), two adapted regularization parameters for land and ocean (right, RMS: 8.08 cm)

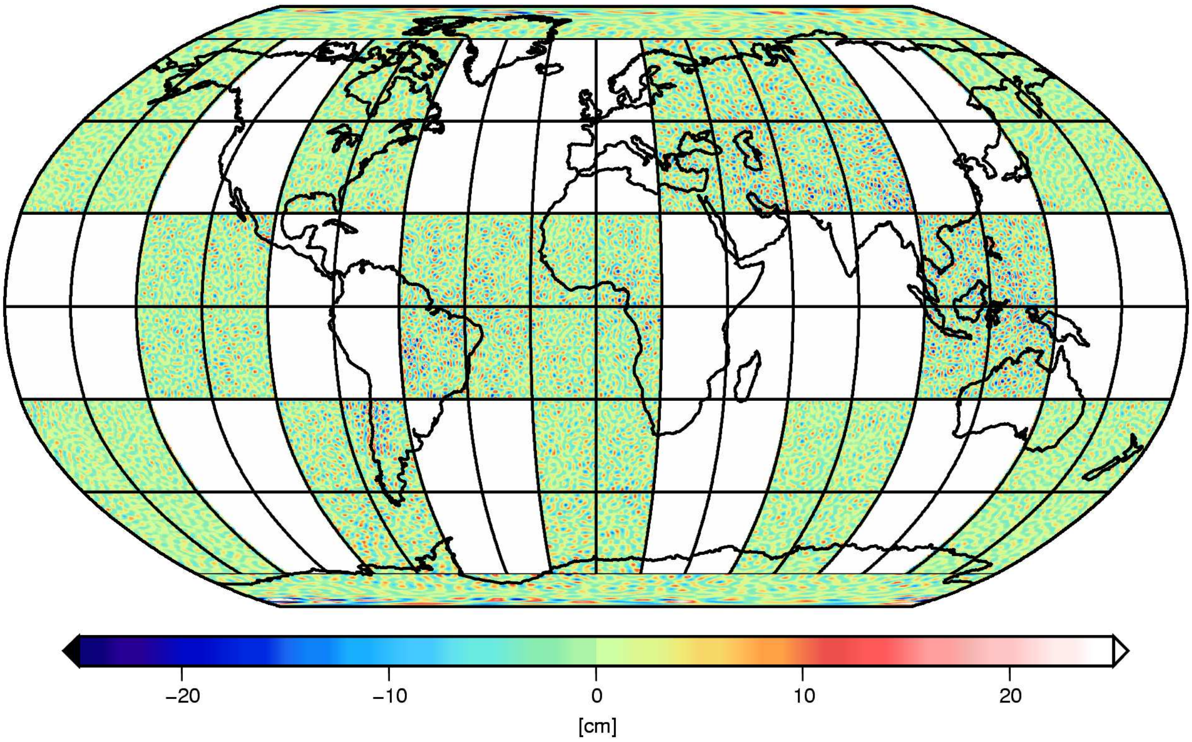


Figure 7.22: Individual patches, differences in geoid heights between combined GRACE/GOCE solution and EGM96

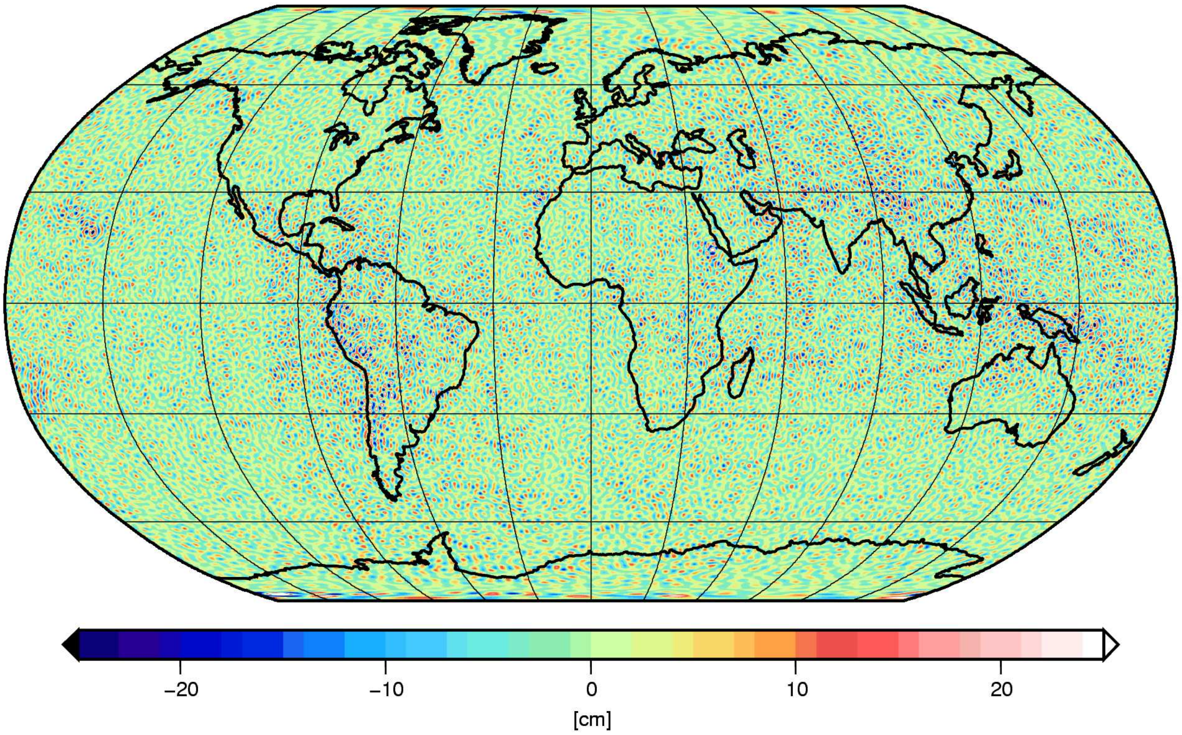


Figure 7.23: Differences in geoid heights between EGM96 and the refined solution, RMS: 6.51 cm

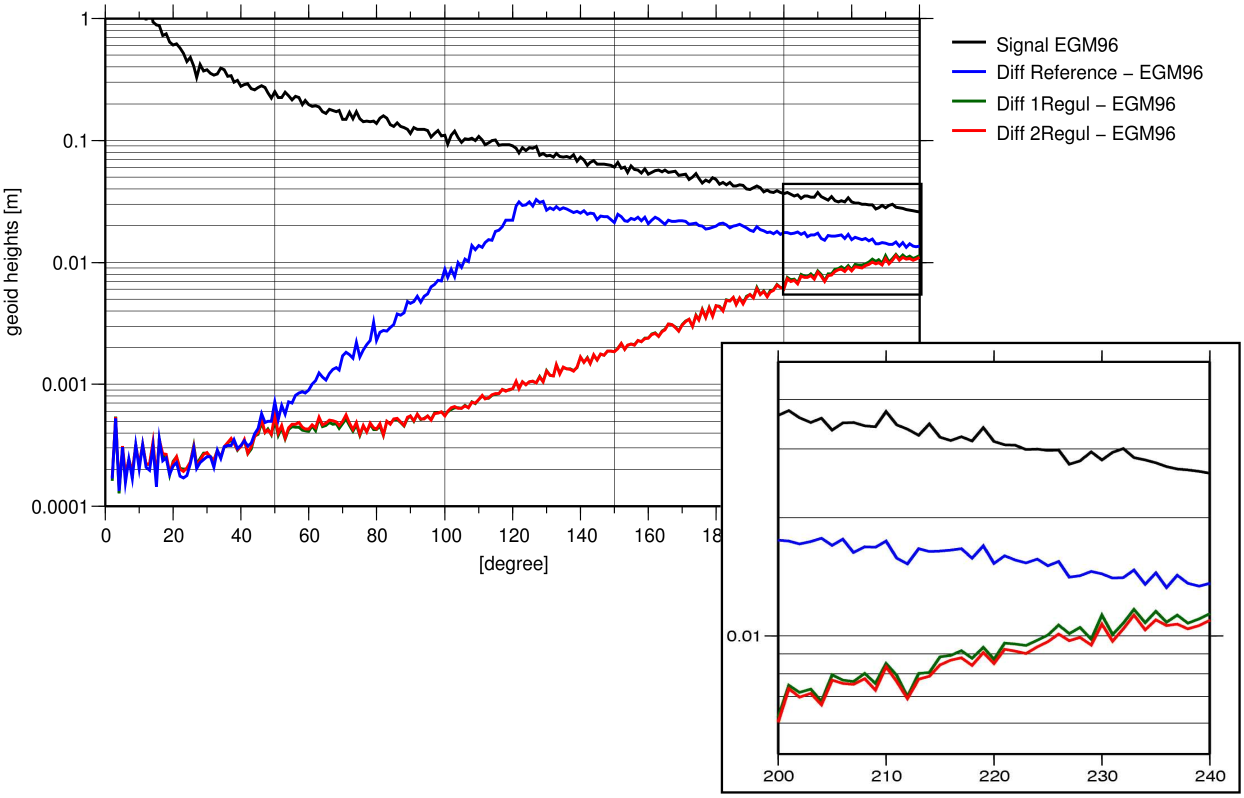


Figure 7.24: Degree variance of the differences between the reference field (GRACE+OSU) and the EGM96 and between the refined solutions and the EGM96

8. Summary and Outlook

In this thesis, an integrated approach was developed for the regional refinement of global gravity field models on the basis of satellite data provided by the missions CHAMP, GRACE, and GOCE. Such an integrated approach comprises the design of specifically tailored space localizing basis functions together with their respective location according to a suitable nodal point pattern, a functional model based on short arcs of the satellite's orbit, and a regional adaptation of the analysis process. Concerning the construction of the basis functions, they were chosen as radial spline kernels, and their shape coefficients were derived from the covariance function of the gravitational potential. In this way, it was achieved that the basis functions reflect the frequency behavior of the gravity field to be modeled. As to the appropriate choice of a spherical grid applied in the location of the spline kernels, several different settings were examined, and a triangular grid obtained from a specific uniform densification of the icosahedron has been proven to be most suitable. It shall be pointed out that, within this thesis, the regional approach does not only imply a regional modeling, but the complete analysis process chain is adapted to explore the benefits of a regional analysis. This starts with choosing only the satellite data originating from over a specific geographical area, resulting in the necessity of using a functional model based on short arcs of the satellite's orbit. The functional model that was described as to the analysis of satellite-to-satellite tracking data is based on the solution of an integral equation of Fredholm type for the high-low mode and on a modification thereof in case of the low-low mode. These models, as well as the pointwise evaluation of the gravity gradient in case of satellite gravity gradiometry, fulfill the requirement of a short-arc analysis and emphasize the in-situ type character of the respective observation principles. Another important aspect in the regional gravity field analysis is the downward continuation process. In this context, especially a regionally adapted regularization was introduced which assigns different regularization matrices to geographical areas with varying signal content. Therefore, a global dampening of the gravity field features can be avoided, as the adapted regularization parameter determined for each region can take into account the varying signal content. Due to this fact, it was expected that it is possible to extract more information from given data than in case of directly calculated global gravity field solutions. To conclude the analysis procedure, an approach was described that combines regional solutions with global coverage to obtain a global solution and to derive the corresponding spherical harmonic coefficients by means of the Gauss-Legendre quadrature method. This combination procedure offers the advantages of a global gravity field solution without losing the details of a regional zoom-in.

The applicability of the regional approach was demonstrated by the analysis of all three satellite missions. In case of CHAMP and GRACE, this implied the analysis of real data, whereas for GOCE, a simulation scenario was presented based on a combined analysis of (simulated) GRACE and GOCE data. Regarding CHAMP, the regional strategy was applied to the refinement of the global CHAMP model ITG-Champ01k, resulting in its significant improvement. The analysis of GRACE was carried out in two steps. In a first step, a solution was calculated from one month of data to demonstrate the performance of the regional procedure. It was possible to show that especially the regionally adapted regularization process can provide a remarkable enhancement in accuracy. The derived regional models showed better agreement with global models derived from a significantly larger amount of data than was the case concerning the respective GFZ-RL04 monthly gravity field. This was true for a comparison with the ITG-Grace02s as well as with the Eigen-GL04C. In a second step, the high resolution model ITG-GraceSpline03 was derived from 4.5 years of GRACE data as a regional refinement of the global model ITG-Grace03s. A considerable gain in resolution was achieved in comparison to the directly calculated global solution, as was illustrated in the space as well as in the frequency domain. Obviously, the regional refinement procedure allowed to extract more information out of the given data set in the high frequencies than was the case for the global model. Concerning the analysis of GOCE, a simulation scenario was designed dealing with the combination of (simulated) GRACE and GOCE data. The target of these investigations was not only to demonstrate the suitability of the regional approach to be used in the GOCE analysis, but also to show its potential to be applied in the combination of different data sets. In the scenario at hand, the GRACE observations provided the information in the long and medium wavelength part in terms of a global reference model, and regional refinements calculated from GOCE measurements enabled the determination of the high resolution gravity field features. Again, the use of the regionally adapted regularization process turned out to be a valuable advantage. In this way, the regional refinement strategy constitutes a flexible combination tool, the method not being limited to the data of the satellite gravity missions, but being suitable for the inclusion of terrestrial, airborne, or altimetry data as well. Especially as these data sets are not available with global coverage, the regional modeling proposes a very adequate possibility of combining them with global satellite models.

Summarizing, it can be concluded that, indeed, a substantial improvement of global models by regional refinement techniques is possible, as was demonstrated by the various results. Especially the signal content in the higher frequency part of the spectrum can be preferably detected using regionally tailored analysis and modeling approaches. Therefore, particular success of the regional approach is expected for the analysis of real GOCE data, hopefully available in the near future, as the primary goal of GOCE is the determination of the short wavelength part of the gravity field. Concerning GOCE, additional advantages of the regional analysis procedure will come into focus, too. On the one hand, this is the ability of dealing with regions with different data coverage which will be interesting when having to cope with the polar gap problem. On the other hand, splitting up the global problem into smaller regional problems provides computational advantages, particularly motivated in case of GOCE by the large amount of data and the large number of projected gravity field parameters.

Despite the satisfactory results already obtained by the regional refinement approach, there is still potential for further improvements of the method. First of all, an advancement of the regional adaptation of the regularization process can be mentioned. The described choice of two regularization areas, one for continental and one for oceanic areas, is a reasonable choice for certain areas, but does not represent an optimal selection for global applications. Further improvement in accuracy could be expected if the regularization areas were more appropriately tailored according to the signal content in different regions. The respective information could be obtained from topographic models or a-priori gravity field solutions. More than two regularization areas per regional solution patch are possible as well.

The possibility of achieving a smoother transition between individually calculated patches and regions with different roughness in the gravity field features should further be investigated. This would prevent the refined solution from remaining truncation effects. These effects are primarily detectable in the very precise determined long wavelength part of the gravitational field, but are too small to have significant influence on the high frequency part, which is the primary focus of the regional refinement approach. Measures that can be taken against the influence of the truncation are larger overlapping regions or the averaging of adjacent solutions in areas close to the border of the patches in the overlapping regions. Thinkable is also to smooth the transition between two adjacent patches (or regularization regions) by a continuous modification of the respective regularization parameters.

So far, the resolution of the spline solutions has been defined globally, and the distance between nodal points has been determined by distributing a specific number of spline kernels as evenly as possible along the nodes of a (global) grid. An advantage of a modeling by space localizing basis functions, however, is the possibility of flexibly adapting the spatial resolution according to the given gravity field characteristics. It would be thinkable, for instance, to locate basis functions with narrow support close together in regions with strong high frequency signal and to use wider kernels located farther apart in areas featuring a rather smooth signal. Furthermore, it might be possible to use the space localizing basis functions to model particular geophysical phenomena. If so, the basis functions could be arranged in a way to best reflect the characteristics of the particular phenomenon, for instance along the course of a river basin, etc.

Furthermore, one could consider multi-resolution modeling techniques as, for example, wavelet representations. Practically speaking, the regional refinement approach with a global (lower resolution) spherical harmonic model and additional regional refinements in terms of radial basis functions represents a two-scale approach. For the given satellite data, the two-scale modeling seems to be well-tailored and sufficient, as proven by the acquired results. But especially when combining the satellite models with terrestrial observations of substantially higher resolution, the extension of the method to further scales and the decomposition of the signal into different detail signals might turn out to be appropriate.

The combination of different regional solutions to one global solution by the use of the Gauss-Legendre quadrature has so far been performed without applying any error propagation; this could be included into the analysis process in the future. However, it has to be considered that the discrete orthogonality of the spherical harmonics, and thus the exactness of the quadrature method, is lost when introducing an additional weight matrix.

Besides the calculation of a static gravity field, GRACE enables the recovery of temporal gravity variations as well. These temporal variations are often regional phenomena, thus a regional modeling seems particularly reasonable. First investigations concerning the regional analysis of time variabilities were performed by EICKER et al. (2007), but are not considered here. The promising results encourage further research in this topic.

A. Mathematical Fundamentals

In this chapter, some basic mathematical concepts are summarized that will be used in the course of the thesis. They can by found, e.g., in BRONSTEIN and SEMENDJAJEW (1995). The first section deals with mathematical (vector) spaces and their characteristics, resulting in the definition of Hilbert spaces. Those are needed, as the construction of basis functions for the gravity field parameterization and their theoretical discussion takes place in Hilbert spaces. In the second section, the concept of linear operators and linear functionals is described.

A.1 Function Spaces

Vector Space A vector space is a set of elements that is closed with respect to the two operations vector addition and scalar multiplication. For each of the two operations certain axioms have to be satisfied.

Norm, Normed Space A normed space X is a vector space with a norm defined on it. The norm $\|\cdot\|$ is a mapping $X \rightarrow \mathbb{R}$ with the following properties for arbitrary $x, y \in X$:

$$\|x\| \geq 0 \tag{A.1}$$

$$\|x\| = 0 \Leftrightarrow x = 0 \tag{A.2}$$

$$\|\alpha x\| = |\alpha| \, \|x\| \tag{A.3}$$

$$\|x + y\| \leq \|x\| + \|y\| \, . \tag{A.4}$$

Cauchy Sequence A sequence in a normed space is called Cauchy sequence if for every positive real number $\epsilon > 0$ there is a positive integer $n \in \mathbb{N}$ such that for all integers $k, l \geq n$ it holds

$$\|u_k - u_l\| < \epsilon. \tag{A.5}$$

Scalar Product, Inner Product If X is a vector space, a scalar product (also called inner product) $\langle \cdot, \cdot \rangle$ is a mapping $X \times X \rightarrow \mathbb{R}$ such that for every $x, y, z \in X$ and $\alpha \in \mathbb{R}$ the following conditions are fulfilled:

$$\langle x, x \rangle \geq 0 \qquad \text{Non-negativity} \tag{A.6}$$

$$\langle x, y \rangle = \langle y, x \rangle \qquad \text{Commutative law} \tag{A.7}$$

$$\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \qquad \text{Linearity} \tag{A.8}$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \, . \qquad \text{Linearity} \tag{A.9}$$

By $\|x\| = \langle x, x \rangle^{\frac{1}{2}}$ the inner product induces a norm $\|\cdot\|$ in X.

Pre-Hilbert Space A normed vector space on which a norm is defined by an inner product is called pre-Hilbert space. If a Pre-Hilbert space is complete with respect to this inner product, it becomes a Hilbert space as described below.

Euclidean Space The n-dimensional vector space $\mathbb{R}^n$ with an inner product $\langle x, y \rangle = \sum\limits_k x_k y_k$ is called Euclidean space.

Completeness A normed space X is called complete if every Cauchy sequence has a limit in X.

Closure Let M be a subset of a normed space X. Then every point $m_0 \in X$ that is a limit of a (convergent) sequence in M is called accumulation point. m_0 itself does not necessarily belong to M. The closure $\overline{M}$ of the subset M includes all points of M together with all accumulation points.

The subset M is closed if every convergent sequence in M has its limit in M.

Complete Subspace A subspace M of a normed space X is complete in itself if, and only if, it is closed in X.

Banach Space Banach spaces are defined as complete normed vector spaces. This means that a Banach space is a vector space V over the real or complex numbers with a norm $\|\cdot\|$ such that every Cauchy sequence in V has a limit in V. In a Banach space the norm does not necessarily have to be defined on the basis of an inner product.

Hilbert Space A Hilbert space $\mathcal{H}$ is an inner product space that is complete under its norm

$$\|x\| = \langle x, x\rangle^{1/2}. \tag{A.10}$$

Thus the norm is defined as the square root of the inner product. The Hilbert space is a generalization of the Euclidean space applied to infinite dimensional spaces such as function spaces. In this way simple geometric concepts such as projection and change of basis can be applied to infinite dimensional spaces. In an equivalent definition, a Hilbert Space can be introduced as an inner product space that is also a Banach space (a complete normed space) under the norm defined by the inner product. In this way, every Hilbert space is a Banach space, because, by definition, a Hilbert space is complete with respect to the norm associated with its inner product, where a norm and an inner product are said to be associated if $\| v \|^2 = (\mathbf{v}, \mathbf{v})$ for all $\mathbf{v}$. The converse is not always true; not every Banach space is a Hilbert space. A necessary and sufficient condition for a Banach space V to be associated to an inner product (which will then necessarily make V into a Hilbert space) is the parallelogram identity,

$$\|u + v\|^2 + \|u - v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right). \tag{A.11}$$

A Hilbert space is *separable* if there exists a countable dense subset S of $\mathcal{H}$, i.e. each element of $\mathcal{H}$ is the limit of a sequence of elements in S, or in other words, if there is a set S with a countable number of elements whose closure is equal to $\mathcal{H}$. A separable Hilbert space possesses a countable orthonormal basis with the basis functions denoted by ϕ_i. If this is the case, each function f in $\mathcal{H}$ can be expressed by an expansion of these basis functions,

$$f = \sum_{i=0}^{\infty} \langle f, \phi_i\rangle \phi_i = \sum_{i=0}^{\infty} f_i \cdot \phi_i, \tag{A.12}$$

where $f_i = \langle f, \phi_i\rangle$ are the coefficients. Parseval's equality states that

$$\|f\|^2 = \langle f, f\rangle = \sum_{i=0}^{\infty} |\langle f, \phi_i\rangle|^2 = \sum_{i=0}^{\infty} f_i^2. \tag{A.13}$$

This means that the norm of the function can be calculated from the sum of the squared coefficients f_i. This sum can be interpreted as the power of a function in the frequency domain. This leads to the interpretation of Parseval's theorem that the power of a function in space domain (given by the norm) equals the power of the same function in the frequency domain.

In the space of square integrable functions $\mathcal{L}_2$, the scalar product is defined by

$$\langle f, g\rangle_{\mathcal{L}_2} = \int f(x)g(x)dx. \tag{A.14}$$

Again Parseval's equality can be used to express the scalar product and the norm as sum of the coefficients of an expansion into basis functions being orthonormal with respect to the scalar product established by Eq. (A.14),

$$\langle f, g \rangle_{\mathcal{L}_2} = \sum_{i=0}^{\infty} f_i \cdot g_i, \tag{A.15}$$

and the norm of a function is given by

$$\|f\|^2_{\mathcal{L}_2} = \langle f, f \rangle_{\mathcal{L}_2} = \sum_{i=0}^{\infty} f_i^2 \qquad < \infty \qquad \Rightarrow \lim_{i \to \infty} f_i = 0. \tag{A.16}$$

Interpreted in the frequency domain, for the function f to be a member of $\mathcal{L}_2$, the norm has to be finite. Thus the coefficients f_i have to approach zero sufficiently fast. In the Section 3.2 about Hilbert spaces with reproducing kernels, it is demonstrated that in those Hilbert spaces the scalar product and thus also the norm of a function are defined in a slightly different way. This norm results in stronger limitations being imposed on the smoothness of the function f and thus on the decreasing behavior of the coefficients. Within the course of this thesis, the scalar product denoted by $\langle \cdot, \cdot \rangle$ and the norm denoted by $\|\cdot\|$ (without subscript) will refer to the respective calculations with respect to $\mathcal{L}_2$.

Sobolev Space The Sobolev space denoted by $W^{k,p}$ is a normed space of functions obtained by imposing the condition of finite L^p norm for given $p \geq 1$ on a function f and its weak derivatives up to some order k,

$$\| u \|_{W^{k,p}} (\Omega) = \left(\sum_{|\alpha| \leq k} \| \partial^\alpha u \|^p_{L^p(\Omega)} \right)^{1/p} . \tag{A.17}$$

A Sobolev-space is complete with respect to this norm; therefore every Sobolev space is also a Banach space. The case $p = 2$ leads to a Hilbert space described above.

A.2 Linear Functionals and Linear Operators

In this section, certain characteristics of linear operators on Hilbert spaces shall be introduced. Details can, for example, be found in Lanczos (1961) and Kress (1989). A linear operator describes the mapping $\mathcal{A} : X \to Y$ from one Hilbert space X onto another Hilbert space Y. This mapping of elements of the respective Hilbert spaces can be expressed by the operator equation

$$\mathcal{A}x = y, \tag{A.18}$$

with $x \in X$ and $y \in Y$. In case of $\mathcal{A}$ representing an integral operator, the operator equation represents a short notation of

$$y(s) = \int K(s,t)x(t)dt, \tag{A.19}$$

where $K(s, t)$ is called the kernel of the operator $\mathcal{A}$.

Linearity An operator $\mathcal{A} : X \to Y$ is called linear if for all elements $f_1, f_2 \in X$ and for all real numbers α_1, α_2 the following equation holds

$$\mathcal{A}(\alpha_1 f_1 + \alpha_2 f_2) = \alpha_1 \mathcal{A} f_1 + \alpha_2 \mathcal{A} f_2. \tag{A.20}$$

Range, Nullspace The range $\mathcal{R}(\mathcal{A})$ of a linear operator $\mathcal{A} : X \to Y$ is defined by

$$\mathcal{R}(\mathcal{A}) = \{\mathcal{A}x \mid x \in X\} . \tag{A.21}$$

It is the subspace of Y that can be reached by applying the operator $\mathcal{A}$ to elements of X.

The nullspace $\mathcal{N}(\mathcal{A})$ of an operator $\mathcal{A} : X \to Y$ is given by

$$\mathcal{N}(\mathcal{A}) = \{x \in X \mid \mathcal{A}x = 0\} , \tag{A.22}$$

which implies that the nullspace consists of those $x \in X$ that are mapped to zero.

Linear Functionals A linear functional is defined as linear map from a Hilbert space X onto the space of real numbers (or in case of a complex space complex numbers),

$$F : X \to \mathbb{R}. \tag{A.23}$$

Therefore, linear functionals represent a special case of linear operators.

Dual Space The space of all linear functionals of X is called dual space of X.

Boundedness A linear functional is bounded if for every $f \in X$ there is a constant $M > 0 \in R$ with

$$|Ff| \leq M \|f\| . \tag{A.24}$$

Riesz Representation Theorem For every bounded linear functional F on a Hilbert space $\mathcal{X}$, there is a unique element $u \in X$ such that $Ff = \langle u, f\rangle$ for all $f \in \mathcal{X}$. An element u being assigned in this way to a linear functional is called the representer of F. Thus every linear functional applied to a function f can be interpreted as the inner product of the representer of the functional with f.

Adjoint Operator, Self-adjoint Operator The adjoint operator $\mathcal{A}^* : Y \to X$ of a bounded linear operator $\mathcal{A} : X \to Y$ is defined by

$$\langle \mathcal{A}x, y\rangle_Y = \langle x, \mathcal{A}^*y\rangle_X . \tag{A.25}$$

For a bounded linear operator $\mathcal{A}$ and its adjoint operator $\mathcal{A}^*$ the following relationships hold

$$R(\mathcal{A})^\perp = N(\mathcal{A}^*) \qquad \text{and} \qquad N(\mathcal{A})^\perp = \overline{R(\mathcal{A}^*)}, \tag{A.26}$$

with $R(\mathcal{A})^\perp$ being the orthogonal complement of $R(\mathcal{A})$. The operator $\mathcal{A} : X \to X$ is called self-adjoint with respect to the scalar product $\langle \cdot, \cdot\rangle_X$ if $\mathcal{A}^* = \mathcal{A}$. The existence of $\mathcal{A}^*$ can be proven by the Riesz theorem. In case of finite dimensional linear operators (i.e. matrices), the adjoint operator can be identified as the transposed matrix.

Boundedness, Norm A linear operator $\mathcal{A} : X \to Y$ is bounded if there exists a positive number α such that

$$\|\mathcal{A}f\|_Y \leq \alpha \|f\|_X \tag{A.27}$$

holds for every $f \in X$. Here $\|\cdot\|_X$ and $\|\cdot\|_Y$ are the norms defined on the spaces X and Y, respectively. In this case, the smallest α for which (A.27) is true for all x is called the operator norm of $\mathcal{A}$ and is denoted by $\|\mathcal{A}\|$.

Continuity For a mapping $\mathcal{A} : X \to Y$ to be continuous, this means that for every $\delta > 0$ there exists a $\epsilon > 0$ such that for $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ it holds

$$\|\mathcal{A}\mathbf{x}_1 - \mathcal{A}\mathbf{x}_2\|_Y = \|\mathbf{y}_1 - \mathbf{y}_2\|_Y < \epsilon \qquad \text{for all } \mathbf{x} \text{ satisfying } \|\mathbf{x}_1 - \mathbf{x}_2\|_X < \delta. \tag{A.28}$$

This implies that small changes in x result in small changes in y. Alternatively, it can be insisted that for every sequence $(x_k) \in X$ it has to hold

$$\lim_{k\to\infty} x_k = x \qquad \Rightarrow \qquad \lim_{k\to\infty} \mathcal{A}x_k = \mathcal{A}x. \tag{A.29}$$

In case of linear operators, boundedness and continuity are equivalent requirements, meaning that a linear operator is continuous if, and only if, it is bounded. In a finite dimensional space, every linear operator is bounded and thus continuous.

Compactness A linear operator $\mathcal{A} : X \to Y$ is compact if, and only if, for every bounded sequence $(x_k) \in X$ there exists a subset (x_k) such that the sequence $(\mathcal{A}x_k)$ converges in Y. A compact linear operator is necessarily bounded and therefore continuous. The concept of compactness plays a decisive role in the discussion of inverse problems, as in infinite dimensional spaces for a compact operator $\mathcal{A}$ the inverse operator $\mathcal{A}^{-1}$ cannot be bounded.

A.2.0.1 Spectral Decomposition

When analyzing linear operators, their spectral analysis presents an important tool. By the evaluation of its eigenvalues or its singular values, relevant characteristics of the operator can be investigated. This helps, for example, to understand the concept of inverse or ill-posed problems that are dealt with in Section 5.1. In the following, the spectral decomposition will be introduced in infinite dimensional Hilbert space, as in the examination of reproducing kernel Hilbert spaces (Section 3.2) the concept of spectral decomposition is needed in the continuous version. The following considerations are limited to compact linear operators that have a discrete spectrum and thus closely resemble the behavior of matrices in finite dimensional spaces. When dealing with discrete ill-posed problems, as in Section 5.1.2, the spectral decomposition is applied to finite dimensional matrices, but the concepts described here can easily be transferred to the finite dimensional case.

Eigenvalue decomposition Every compact, self-adjoint, positive (semi) definite linear operator $\mathcal{T} : X \to X$ can be characterized by its eigenvalues and corresponding eigenfunctions. If the following relationship,

$$\mathcal{T}v_i = \lambda_i v_i, \tag{A.30}$$

holds for a $v_i \neq 0$, then λ_i are called eigenvalues of the operator $\mathcal{T}$, and v_i are the eigenfunctions associated with the respective eigenvalues. It can be shown that $\mathcal{A}$ has a finite or countably infinite set of eigenvalues, and their only possible accumulation point is zero. Therefore, they can be arranged in descending order according to

$$\lambda_1 \geq \lambda_2 \geq ... \geq 0. \tag{A.31}$$

This feature is stated in the Hilbert-Schmidt theorem which indicates that for a compact, self-adjoint operator there is a sequence of non-zero eigenvalues such that

$$\lim_{i\to\infty} \lambda_i = 0. \tag{A.32}$$

The set of eigenvalues is called the spectrum of the operator. The norm of $\mathcal{T}$ is equal to the largest eigenvalue,

$$\|\mathcal{T}\| = \lambda_1, \tag{A.33}$$

and it is also called spectral radius of $\mathcal{T}$. This definition of the norm of an operator is equivalent to the definition given by (A.27). The eigenfunctions are orthogonal and form a basis of $\overline{R(\mathcal{T})}$, i.e. of the closure of the range of $\mathcal{T}$. $R(\mathcal{T})$ is not necessarily complete, therefore its closure is required. The mapping $\mathcal{T} : X \rightarrow X$ can then be expressed by the following spectral expansion,

$$\mathcal{T}x = \sum_{i=1}^{\infty} \lambda_i \langle x, v_i \rangle v_i. \tag{A.34}$$

Singular Value Decomposition The operators $\mathcal{A}^*\mathcal{A} : X \rightarrow X$ and $\mathcal{A}\mathcal{A}^* : Y \rightarrow Y$, with $\mathcal{A}$ according to Eq. (A.18), are compact and self-adjoint under the assumption that $\mathcal{A}$ is compact. Both operators have the same eigenvalues, and the following equations can be formulated in equivalence with Eq. (A.30),

$$\begin{aligned} \mathcal{A}^*\mathcal{A}v_i &= \lambda_i v_i \\ \mathcal{A}\mathcal{A}^*u_i &= \lambda_i u_i. \end{aligned} \tag{A.35}$$

By these relations, the singular value expansion of the operator $\mathcal{A} : X \rightarrow Y$ can be derived. With u_i being the eigenfunctions of $\mathcal{A}\mathcal{A}^*$ and v_i being the eigenfunctions of $\mathcal{A}^*\mathcal{A}$, it follows with $\sigma_i = \sqrt{\lambda_i}$,

$$\begin{aligned} \mathcal{A}v_i &= \sigma_i u_i \\ \mathcal{A}^*u_i &= \sigma_i v_i. \end{aligned} \tag{A.36}$$

The triplet $\{\sigma_i, u_i, v_i\}$ is called the singular system of $\mathcal{A}$, with the singular values σ_i and the left and right singular functions u_i and v_i, respectively. Eq. (A.36) shows that by the operator $\mathcal{A}$ any singular function v_i is mapped onto the corresponding function u_i, with the singular value σ_i describing the scaling of the mapping. The singular values have zero as accumulation point and can be ordered in decreasing order, as is the case with the eigenvalues,

$$\sigma_1 \geq \sigma_2 \geq ... \geq 0. \tag{A.37}$$

The singular functions u_i are a complete orthogonal system (thus a basis) for $\overline{R(\mathcal{A})}$, and the singular functions v_i constitute a basis for $\overline{R(\mathcal{A}^*)}$. The mapping $\mathcal{A}x$ of a function x onto $R(\mathcal{A})$ can be described by

$$\mathcal{A}x = \sum_{i=1}^{\infty} \sigma_i \langle x, v_i \rangle u_i. \tag{A.38}$$

Here the smoothing effect of the operator becomes evident. As is shown in Chapter 5.1.1 for operators relevant in this thesis, small singular values can be associated with high frequencies. As the singular values σ_i decay with increasing i, those high frequency components in x are dampened during the mapping process. On the other hand, when solving for x from the operator equation Eq. (A.18), the solution has the following spectral expansion,

$$\hat{x} = \sum_{i=1}^{\infty} \frac{\langle y, u_i \rangle}{\sigma_i} u_i. \tag{A.39}$$

Here the higher frequencies become amplified due to division by small singular values. For the solution Eq. (A.39) to exist, its norm has to be finite, which results in the following requirement,

$$\sum_{i=1}^{\infty} \left(\frac{\langle y, u_i \rangle}{\sigma_i} \right)^2 < \infty. \tag{A.40}$$

This is known as Picard condition, and it implies a smoothness requirement on the observations y, as the coefficients $\langle y, u_i \rangle$ have to decay fast enough compared to the singular values to fulfill Eq. (A.40). Otherwise the expansion (A.39) does not converge. This requirement is equivalent to the demand that y has to belong to $R(\mathcal{A})$.

B. Derivation of the Weights for the Gauss-Legendre Quadrature

The special feature of the Gauss-Legendre quadrature is the fact that polynomials up to an order of $2N-1$ can be integrated exactly, even though only N evaluation points are used. This requires the quadrature nodes and weights to be chosen appropriately. It is the goal to determine quadrature weights w_i that fulfill the following equation,

$$\int\limits_a^b f(t)dt = \sum_{i=1}^{N} f(t_i)\bar{w}_i. \tag{B.1}$$

Let $f(t)$ be a polynomial of degree $2N-1$ and $P_n(t)$ be a set of orthogonal polynomials, in case of the Gauss-Legendre quadrature these are the Legendre polynomials. Then for P_N being the Legendre polynomial of degree N, there are polynomials q_{N-1} and r_{N-1}, each of degree $N-1$, that fulfill

$$f(t) = r_{N-1}(t) + P_N(t)\, q_{N-1}(t). \tag{B.2}$$

This can immediately be understood for the case of $p_N(t) = t^N$,

$$f(t) = f_0 + f_1 t + \ldots + f_{2N-1}t^{2N-1} = f_0 + \ldots + f_{N-1}t^{N-1} + t^N(f_N + f_{N+1}t + \ldots + f_{2N-1}t^{N-1}). \tag{B.3}$$

Integrating both sides of Eq. (B.2) yields

$$\int_{-1}^{1} f(t)dt = \int_{-1}^{1} r_{N-1}(t)\, dt + \int_{-1}^{1} P_N(t)\, q_{N-1}(t)dt = \int_{-1}^{1} r_{N-1}(t)dt, \tag{B.4}$$

as q_{N-1} can be expressed as a linear combination of $P_0, \ldots, P_{N-1}$ and is, therefore, orthogonal to $P_N(x)$ which results in the second integral on the right side to vanish. The evaluation points t_i are chosen at the zeros of the Legendre polynomial P_N; therefore from Eq. (B.2) it follows $f(t_i) = r(t_i)$ due to $P_N(t_i) = 0$. For the following procedure the Lagrange polynomials

$$L_i(t) = \prod_{l=1, l\neq i}^{N} \frac{t - t_l}{t_i - t_l} \tag{B.5}$$

have to be introduced. They are equal to one at $t = t_i$ and zero at the other points t_l for $l \neq i$. That is why they are denominated as interpolatory functions. $r_{N-1}(t)$ is of degree $N-1$, therfore it can be expressed as a linear combination of the Lagrange polynomials according to

$$r_{N-1}(t) = \sum_{i=1}^{N} f(t_i)L_i(t). \tag{B.6}$$

Inserting this into (B.4) and a comparison with (6.18) provides

$$\int_{-1}^{1} f(t)dt = \int_{-1}^{1} \sum_{i=1}^{N} f(t_i)L_i(t)\, dt = \sum_{i=1}^{N} f(t_i) \underbrace{\int_{-1}^{1} L_i(t)\, dt}_{\bar{w}_i}. \tag{B.7}$$

The Lagrange polynomial can be formulated in terms of the Legendre polynomials according to

$$L_i(t) = \frac{P_N(t)}{(t - t_i)P'_N(t_i)} \tag{B.8}$$

to simplify the calculation of the weights. Eq. (B.8) can be understood when taking into account that the evaluation points are located at $P_N(t_i) = 0$, and thus the right side of Eq. (B.8) has the same $N-1$ zeros as the right side of Eq. (B.5). Furthermore, both fractions are equal to one in case of $t = t_i$. Thus both phrases for L_i have N coinciding values, and since a polynomial of degree $N-1$ is sufficiently determined by N specific values, both expressions can be regarded as equal. This leads to the weights for the Gauss-Legendre quadrature method,

$$\bar{w}_i = \int_{-1}^{1} \frac{P_N(t)}{(t-t_i)P_N'(t_i)} dt, \tag{B.9}$$

or in an alternative representation,

$$\bar{w}_i = \frac{A_N}{A_{N-1}} \int_{-1}^{1} \frac{[P_{N-1}(t)]^2}{P_{N-1}(t_i)P_N'(t_i)} dt. \tag{B.10}$$

Here P_L' refers to the first derivative of the Legendre polynomial of degree L, and A_n is the coefficient of t^n in P_n. It can be calculated by

$$A_n = \frac{(2n)!}{2^n (n!)^2}, \tag{B.11}$$

which follows from the closed representation of the Legendre polynomials by the sum

$$P_n(t) = \sum_{k=0}^{N} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} t^{n-2k}. \tag{B.12}$$

From (B.11) it can thus be concluded

$$\frac{A_N}{A_{N-1}} = \frac{2N-1}{N}. \tag{B.13}$$

With the simplification

$$\int_{-1}^{1} [P_n(t)]^2 dt = \frac{2}{2n+1} \qquad \Rightarrow \qquad \int_{-1}^{1} [P_{N-1}(t)]^2 dt = \frac{2}{2N-1}, \tag{B.14}$$

Eq. (B.10) can be rewritten according to

$$\bar{w}_i = \frac{2N-1}{N} \cdot \frac{2}{2N-1} \cdot \frac{1}{P_{N-1}(t_i)P_N'(t_i)} = \frac{2}{NP_{N-1}(t_i)P_N'(t_i)}. \tag{B.15}$$

As a last step, the recurrence relation

$$(1-t^2)P_n'(t) = nP_{n-1}(t) - ntP_n(t) \qquad \Rightarrow \qquad (1-t_i^2)P_N'(t_i) = NP_{N-1}(t_i) \tag{B.16}$$

can be applied leading to the following version of the quadrature weights that has been applied in the calculations described in this thesis:

$$\bar{w}_i = \frac{2}{(1-t_i^2)(P_L'(\cos(\theta_i)))^2}. \tag{B.17}$$

It has to be pointed out that $\bar{w}_i$ only denotes the weights for the one-dimensional quadrature along the meridians. The quadrature weights to be used in Eq. (6.6) also include the longitude dependent part and can be calculated by

$$w_i = \Delta\lambda \bar{w}_i. \tag{B.18}$$

Abbreviations

CHAMP	Challenging Minisatellite Payload
CSR	Center of Space Research
DEOS	Department of Earth Observation and Space Systems, TU Delft
DLR	Deutsches Zentrum für Luft- und Raumfahrt
EGM	Earth Gravitational Model
EIGEN	European Improved Gravity Model of the Earth by New Techniques
ESA	European Space Agency
FFT	Fast Fourier Techniques
GFZ	GeoForschungsZentrum Potsdam
GGM	GRACE Gravity Model
GMST	Greenwich Mean Sidereal Time
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer
GPS	Global Positioning System
GRACE	Gravity Recovery And Climate Experiment
GRF	Gradiometer Reference Frame
GROOPS	Gravity Field Object Oriented Programming System
ICRF	International Celestial Reference Frame
IERS	International Earth Rotation and Reference Systems Service
ITG	Institut fuer Theoretische Geodäsie, Universität Bonn
ITRF	International Terrestrial Reference Frame
JPL	Jet Propulsion Laboratory
KBR	K-Band Ranging System
LEO	Low Earth Orbiter
LOS	Line of Sight
NASA	National Aeronautics and Space Administration
POD	Precise Orbit Determination
PSD	Power Spectral Density
RKHS	Reproducing Kernel Hilbert Space
RMS	Root Mean Square (Error)
SDS	Science Data System
SLR	Satellite Laser Ranging
SST	Satellite-to-Satellite Tracking
VCE	Variance Component Estimation
TUM	Technische Universität München
UTCSR	University of Texas in Austin, Center for Space Research
XML	Extensible Markup Language

List of Figures

List of Tables

References

Aronszajn, N. (1950) Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68(3):337–404.

Austen, G. and T. Reubelt (2000) Räumliche Schwerefeldanalyse aus semi-kontinuierlichen Ephemeriden niedrigfliegender GPS-vermessener Satelliten vom Typ CHAMP, GRACE und GOCE. Diplomarbeit, Universität Stuttgart.

Balmino, G. (1972) Representation of the earth potential by buried masses. In: Henriksen, S.W., A. Mancini and B.H. Chovitz (Eds.), *Use of Artificial Satellites for Geodesy and Geodynamics.* American Geophysical Union, Washington, 121–124.

Barthelmes, F. (1986) Untersuchungen zur Approximation des äußeren Gravitationsfeldes der Erde durch Punktmassen mit optimierten Positionen. *Veröffentlichungen des Zentralinstituts für Physik der Erde, Potsdam*, 72.

Baur, O. and N. Sneeuw (2007) Slepian approach revisited: new studies to overcome the polar gap. In: *Proceedings of the 3rd GOCE User Workshop, Frascati, Italy.*

Björck, A. (1996) *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia.

Bouman, J. (1998) *Quality of Regularization Methods.* DEOS Report No. 98.2. Delft University Press.

Bronstein, I.N. and K.A. Semendjajew (1995) *Taschenbuch der Mathematik.* B. G. Teubner, Leipzig.

Colombo, O.L. (1981) *Numerical methods for harmonic analysis on the sphere.* Number 310 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

Conway, J.H. and N.J.A. Sloane (1998) *Sphere Packings: Lattices and Groups.* Springer.

Delaunay, B. (1934) Sur la sphère vide. *Bulletin of Academy of Sciences of the USSR 7*, 6:793–800.

Ditmar, P. and A.A. van Eck van der Sluijs (2004) A technique for modeling the Earth's gravity field on the basis of satellite accelerations. *Journal of Geodesy*, 78(1–2):12–33.

Ditmar, P., R. Klees and F. Kostenko (2003) Fast and accurate computation of spherical harmonic coefficients from satellite gravity gradiometry data. *Journal of Geodesy*, 76(11):690–705.

Ditmar, P., V. Kuznetsov, A.A. van Eck van der Sluijs, E. Schrama and R. Klees (2006) "DEOS-CHAMP-01C-70": a model of the Earth's gravity field computed from accelerations of the CHAMP satellite. *Journal of Geodesy*, 79:586–601.

Drinkwater, M.R., R. Floberhagen, R. Haagmans, D. Muzi and A. Popescu (2003) GOCE: ESA's first Earth Explorer Core mission. *Space Science Series of ISSI*, 18:419–432.

Driscoll, J.R. and D.M. Healy (1994) Computing Fourier transforms and convolutions on the 2-sphere. *Adv. Appl. Math.*, 15(2):202–250.

Edmundson, J.R. (1992) The distribution of charged particles on the surface of a sphere. *Acta Cryst.*, A48:60–69.

Eicker, A., T. Mayer-Gürr and K.H. Ilk (2004) A global CHAMP gravity field by merging of regional refinement patches. In: *Proceedings of the Joint CHAMP/GRACE Science Meeting.* www.gfz-potsdam.de/pb1/JCG/Eicker-etal_jcg.pdf.

EICKER, A., T. MAYER-GÜRR and K.H. ILK (2005) Global gravity field solutions based on a simulation scenario of GRACE SST data and regional refinements by GOCE SGG observations. In: JEKELI, C., L. BASTOS and J. FERNANDES (Eds.), *Gravity Geoid and Space Mission, GGSM 2004*, Volume 129 of the series International Association of Geodesy Symposia. Springer, 66–71.

EICKER, A., T. MAYER-GÜRR and K.H. ILK (2006) An integrated global/regional gravity field determination approach based on GOCE observations. RUMMEL, R., CH. REIGBER, M. ROTHACHER, G. BOEDECKER, U. SCHREIBER and J. FLURY (Eds.), *Observation of the Earth System from Space*, Springer, Berlin - Heidelberg, 225–237.

EICKER, A., T. MAYER-GÜRR and K.H. ILK (2007) Improved resolution of a GRACE gravity field model by regional refinements. In: *submitted to Proceedings IUGG General Assembly, Perugia.*

ESA (1999) *Gravity Field and Steady–State Ocean Circulation Mission.* Reports for Mission Selection, ESA SP–1233(1). European Space Agency Publications Division, Noordwijk.

FENGLER, M.J., W. FREEDEN and V. MICHEL (2004) The Kaiserslautern multiscale geopotential model SWITCH-03 from orbit perturbations of the satellite CHAMP and its comparison to the models EGM96, UCPH2002_02_0.5, EIGEN-1s and EIGEN-2. *Geophysical Journal International*, 157:499–514.

FENGLER, M.J., W. FREEDEN, A. KOHLHAAS, V. MICHEL and T. PETERS (2007) Wavelet modelling of regional and temporal variations of the Earth's gravitational potential observed by GRACE. *Journal of Geodesy*, 81:5–15.

FLECHTNER, F. (2005) AOD1B Product Description Document. Technical Report GRACE 327-750, Jet Propulsion Labratory. http://podaac.jpl.nasa.gov/grace/documentation.html.

FLECHTNER, F. (2007) Release notes for GFZ RL04 GRACE L2 products. http://isdc.gfz-potsdam.de/index.php?name=UpDownload&req=getit&lid=427.

FÖLDVARY, L., D. ŠVEHLA, C. GERLACH, T. GRUBER, M. WERMUTH, N. SNEEUW, B. FROMMKNECHT, H. OBERNDORFER, T. PETERS, M. ROTHACHER, R. RUMMEL and P. STEINBERGER (2005) Gravity model TUM-2Sp based on the energy balance approach and kinematic CHAMP orbits. In: REIGBER, C., H. LUEHR, P. SCHWINTZER and J. WICKERT (Eds.), *Earth observation with CHAMP – Results from three years in orbit.* Springer, 13–18.

FÖRSTE, C., F. FLECHTNER, R. SCHMIDT, U. MEYER, R. STUBENVOLL, F. BARTHELMES, R. KÖNIG, K.H. NEUMAYER, M. ROTHACHER, CH. REIGBER, R. BIANCALE, S. BRUINSMA, J.-M. LEMOINE and J.C. RAIMONDO (2005) A new high resolution global gravity field model derived from combination of GRACE and CHAMP mission and altimetry/gravimetry surface gravity data. *Poster presented at EGU General Assembly 2005, 24–29 April, Vienna, Austria.*

FREEDEN, W. (1981) On spherical spline interpolation and approximation. *Math. Meth. in the Appl. Sci.*, 3:551–575.

FREEDEN, W. (1999) *Multiscale Modelling of Spaceborne Geodata.* B.G. Teubner, Stuttgart.

FREEDEN, W., T. GERVENS and M. SCHREINER (1998) *Constructive Approximation on the Sphere.* Oxford University Press, Oxford.

GERLACH, C., L. FÖLDVARY, D. ŠVEHLA, T. GRUBER, M. WERMUTH, N. SNEEUW, B. FROMMKNECHT, H. OBERNDORFER, T. PETERS, M. ROTHACHER, R. RUMMEL and P. STEIGENBERGER (2003) A CHAMP-only gravity field model from kinematic orbits using the energy integral. *Geophysical Research Letters*, 30(20).

GFZ POTSDAM (2007) Gravity Field Model Eigen-GL04c. http://www.gfz-potsdam.de/pb1/op/grace/results/grav/g005_eigen-gl04c.html.

GROETSCH, C.W. (1977) *Generalized Inverses of Linear Operators.* Dekker, New York.

HADAMARD, J. (1923) *Lectures on the Cauchy Problem in Linear Partial Differential Equations.* Yale University Press, New Haven.

HAJELA, D.P. (1973) *The Computation of* 15° *and* 10° *Equal Area Block Terrestrial Free Air Gravity Anomalies.* Number 194 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

HAN, S.C. and F.J. SIMONS (2008) Spatiospectral localization of global geopotential fields from GRACE reveals the coseismic gravity change due to the 2004 Sumatra-Andaman earthquake. *Journal of Geophysical Research*, in press.

HANSEN, P.C. (1987) The truncated SVD as a method for regularization. *BIT Computer Science and Numerical Mathematics*, 27:534–553.

HANSEN, P.C. (1992) Analysis of discrete ill–posed problems by means of the L–curve. *SIAM Review*, 34(4):561–580.

HANSEN, P.C. (1997) *Rank–Deficient and Discrete Ill–Posed Problems.* SIAM, Philadelphia.

HEISKANEN, W.A. and H. MORITZ (1967) *Physical Geodesy.* Freeman, San Francisco.

HEITZ, S. and E. STÖCKER-MEIER (1994) *Grundlagen der Physikalischen Geodäsie.* Dümmler, Bonn.

HOBSON, E.M. (1931) *The Theory of Spherical and Ellipsoidal Harmonics.* Cambridge University Press, Cambridge.

HOLSCHNEIDER, M. (1995) *Wavelets: An Analysis Tool.* Oxford University Press, Oxford.

HOWE, E., L. STENSENG and C.C. TSCHERNING (2003) Analysis of one month of CHAMP state vector and accelerometer data for the recovery of the gravity potential. *Advances in Geosciences*, 1:1–4.

ILK, K.H. (1984) On the analysis of satellite–to–satellite tracking data. In: *Proceedings of the International Symposium on Space Techniques for Geodesy.* Sopron, Hungary, 59–64.

ILK, K.H., A. LÖCHER and T. MAYER-GÜRR (2005A) Do we need new gravity field recovery techniques for the new gravity field satellites? In: *Proceedings of the IV Hotine-Marussi Symposium of Theoretical and Computational Geodesy: Challenge and Role of Modern Geodesy, May 29 - June 2.* Wuhan, China.

ILK, K.H., J. FLURY, R. RUMMEL, P. SCHWINTZER, W. BOSCH, C. HAAS, J. SCHRÖTER, D. STAMMER, W. ZAHEL, H. MILLER, R. DIETRICH, P. HUYBRECHTS, H. SCHMELING, D. WOLF, J. RIEGGER, A. BARDOSSY and A. GÜNTNER (2005B) *Mass Transport and Mass Distribution in the Earth System – Contribution of the New Generation of Satellite Gravity and Altimetry Missions to Geosciences, Proposal for a German Priority Research Program.* GOCE-Projektbüro Deutschland, Technische Universität München, GeoForschungsZentrum Potsdam, 2. Edition.

ILK, K.H., A. EICKER and T. MAYER-GÜRR (2006) Global gravity field recovery by merging regional focusing patches: an integrated approach. In: RIZOS, C. (Eds.), *Proceedings of the IAG International Symposium Dynamic Planet.* Cairns, Australia.

JÄGGI, A., G. BEUTLER and R. DACH (2007) Assessment of GPS observables for gravity field recovery from GRACE. *Proceedings of the IUGG General Assembly, Perugia (submitted).*

JEKELI, C. (1999) The determination of gravitational potential differences from satellite-to-satellite tracking. *Celestial Mechanics and Dynamical Astronomy*, 7582:85–100.

JEKELI, CHRISTOPHER (2005) *Spline Representation of Functions on a Sphere for Geopotential Modeling.* Number 475 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

KAULA, W.M. (1966) *Theory of Satellite Geodesy.* Blaisdell, Waltham.

Keller, W. (Eds.) (2004) *Wavelets in Geodesy and Geodynamics.* Walter de Gruyter, Berlin.

Kenner, H. (1976) *Geodesic Math and How to Use It.* University of California Press, Berkely.

Koch, K.R. (1990) *Bayesian Inference with Geodetic Applications.* Springer, Berlin.

Koch, K.R. (1997) *Parameterschätzung und Hypothesentests.* Dümmler, Bonn.

Koch, K.R. and J. Kusche (1907) Comments on Xu et al. (2006) Variance component estimation in linear inverse ill-posed models, J Geod 80(1):69-81. *Journal of Geodesy*, 81:629–631.

Koch, K.R. and J. Kusche (2001) Regularization of geopotential determination from satellite data by variance components. *Journal of Geodesy*, 76(5):641 – 652.

Koop, R. (1993) *Global Gravity Field Modelling using Satellite Gravity Gradiometry.* Publications on Geodesy. New Series no. 38. Netherlands Geodetic Commission, Delft.

Kottwitz, D.A. (1991) The densest packing of equal circles on a sphere. *Acta Cryst.*, A4-7.

Krarup, T. (1969) A Contribution to the Mathematical Foundation of Physical Geodesy. *Dan. Geod. Inst. Copenhagen*, 44.

Kress, R. (1989) *Linear Integral Equations.* Applied Mathematical Sciences 82. Springer, Berlin.

Kusche, J. (2002) *Inverse Probleme bei der Gravitationsfeldbestimmung mittels SST- und SGG-Satellitenmissionen.* Number 548 in Reihe C, Deutsche geodätische Kommission bei der Bayerischen Akademie der Wissenschaften. C.H. Beck'sche Verlagsbuchhandlung, München.

Kusche, J. and J. P. van Loon (2004) Statistical assessment of CHAMP data and models using the energy balance approach. In: Reigber, C., H. Luehr, P. Schwintzer and J. Wickert (Eds.), *Earth observation with CHAMP.* Springer.

Lanczos, C (1956) *Applied Analysis.* Prentice Hall, Englewood Cliffs, N.J.

Lanczos, C (1961) *Linear Differential Operators.* Van Nostrand, London.

Landau, H.J. and H.O. Pollak (1962) Prolate spheroidal wave functions, Fourier analysis and uncertainty-III: The dimension of the space of essentially time- and band-limited signals. *Bell Syst. Tech. J.*, 41:1295–1336.

Landweber, L. (1951) An iterative formula for Fredholm integral equations of the first kind. *Am. J. Math.*, 73:615–624.

Laprise, R. (1992) The resolution of global spectral models. *Bulletin American Meteorological Society*, 73:1453–1454.

Le Provost, C. (2001) Ocean tides. In: Fu, L.L. and A. Cazenave (Eds.), *Satellite Altimetry and Earth Sciences.* Springer, 267–303.

Lelgemann, D. (1981) On numerical properties of interpolation with harmonic kernel functions. *Manuscripta Geodaetica*, 6:157–191.

Lemoine, F.G., S.C. Kenyon, J.K. Factor, R.G. Trimmer, D.S. Pavlis, N.K. Chinn, C.M. Cox, S.M. Klosko, S.B. Luthcke, M.H. Torrence, Y.M. Wang, R.G. Williamson, E. C. Pavlis, R.H. Rapp and T.R. Olson (1998) *The Development of the Joint NASA GSFC and NIMA Geopotential Model EGM96.* NASA Goddard Space Flight Cent., Greenbelt, Md.

Löcher, A. and K.H. Ilk (2005) Energy balance relations for validation of gravity field models and orbit determinations applied to the results of the CHAMP mission. In: Reigber, C., H. LÃ¼hr, P. Schwintzer and J. Wickert (Eds.), *Earth observation with CHAMP – Results from three years in orbit.* Springer, 53–58.

LÖCHER, A. and K.H. ILK (2006) A validation procedure for satellite orbits and force function models based on a new balance equation approach. In: RIZOS, C. (Eds.), *Proceedings of the IAG International Symposium Dynamic Planet.* Cairns, Australia.

LOUIS, A.K. (1989) *Inverse und schlecht gestellte Probleme.* Teubner, Stuttgart.

MAYER-GÜRR, T. (2006) *Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE.* Dissertation at the University of Bonn. URN: urn:nbn:de:hbz:5N-09047, URL: http://hss.ulb.uni-bonn.de/diss_online/landw_fak/2006/mayer-guerr_torsten.

MAYER-GÜRR, T., K.H. ILK, A. EICKER and M. FEUCHTINGER (2005) ITG-CHAMP01: a CHAMP gravity field model from short kinematic arcs over a one-year observation period. *Journal of Geodesy,* 78(8):462–480.

MAYER-GÜRR, T., A. EICKER and K.H. ILK (2006) Gravity field recovery from GRACE-SST data of short acrs. FLURY, J., R. RUMMEL, C. REIGBER, G. BOEDECKER M. ROTHACHER and U. SCHREIBER (Eds.), *Observation of the Earth System from Space*, Springer, Berlin - Heidelberg, 131–148.

MAYER-GÜRR, T., A. EICKER and K.H. ILK (2007A) ITG-Grace02s: A GRACE gravity field derived from short arcs of the satellites orbit. In: *Procedings of the 1st International Symposium of the International Gravity Field Sevice "Grvaity Field of the Earth", Istanbul*, 193–198.

MAYER-GÜRR, T., A. EICKER and K.H. ILK (2007B) ITG-Grace03 Gravity Field Model. http://www.geod.uni-bonn.de/itg-grace03.html.

MCCARTHY, D.D. and G. PETIT (Eds.) (2004) *IERS Conventions 2003.* Number 32 in IERS Technical Notes. Verlag des Bundesamts fuer Kartographie und Geodäsie, Frankfurt am Main.

MEISSL, P. (1971) *A Study of Covariance Functions Related to the Earth's Disturbing Potential.* Number 151 in Reports of the Department of Geodetic Science. Ohio State University, Columbus, Ohio.

MERCER, J. (1909) *Functions of positive and negative type and their connection with the theory of integral equations.* Philos. Trans. Roy. Soc., London.

MESCHKOWSKI, H. (1962) *Hilbertsche Räume mit Kernfunktionen.* Die Grundlehren der Mathematischen Wissenschaften 113. Springer, Berlin.

MOHLENKAMP, M. (1997) *A Fast Transform for Spherical Harmonics.* Dissertation Presented to the Faculty of the Graduate School, Yale University.

MORITZ, H. (1962) *Interpolation and Prediction of Gravity and Their Accuracy.* Number 24 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

MORITZ, H. (1966) *Methods for Downward Continuation of Gravity.* Reihe A, Nr. 50. Deutsche Geodätische Kommission, München.

MORITZ, H. (1978) Least-squares collocation. *Reviews of Geophysics and Space Physics*, 16:421–430.

MORITZ, H. (1978) *Statistical Foundations of Collocation.* Number 272 in Reports of the Department of Geodetic Science. Ohio State University, Columbus, Ohio.

MORITZ, H. (1980) *Advanced Physical Geodesy.* Wichmann, Karlsruhe.

MORITZ, H. and SANSÒ (1981) A dialogue on collocation. In: *Proceedings of the Symposium on Mathematical Geodesy (IV Hotine Symposium), Firenze, Italy.* Commissione Geodetica Italiana.

MOROZOV, V.A. (1966) On the solution of functional equations by the method of regularization. *Soviet Math. Dokl*, 7:414–417.

MULLER, P.M. and W.L. SJOGREN (1968) Mascons: lunar mass concentrations. *Science*, 161:680–684.

NEUMANN, F. (1838) Über eine Eigenschaft der Laplaceschen Y^n und ihre Anwendung zur analytischen Darstellung derjenigen Phänomene, welche Functionen der geographischen Länge und Breite sind. *Schumachers Astron. Nachrichten*, 15:313–325.

NIEDERREITER, H. (1992) *Random Number Generation and Quasi-Monte Carlo Method.* SIAM, Philadelphia.

NIEMEIER, W. (2002) *Ausgleichungsrechnung.* Walter de Gruyter, Berlin.

PAVLIS, N.K., S.A. HOLMES, S.C. KENYON, D. SCHMIDT and R. TRIMMER (2005) A preliminary gravitational model to degree 2160. In: JEKELIS, C., L. BASTOS and J. FERNANDES (Eds.), *Gravity, Geoid and Space Missions.* Springer, 18–23.

PAYNE, M.H. (1971) *Truncation effects in geopotential modelling.* Analytical Mechanics Associates, Seabrook, Maryland.

PHILIPS, D. (1962) A technique for the numerical solution of certain intergral equations of the first kind. *Journal of the Association for Computing Machinery*, 9:84–97.

PRANGE, L., A. JÄGGI, G. BEUTLER and R. DACH (2007) Gravity field determination at the AIUB - the celestial mechanic approach. *Proceedings of the IUGG General Assembly, Perugia (submitted).*

RAPP, R.H. (1971) Equal area blocks. *Bulletin Geodesique*, 99.

RAPP, R.H., Y.M. WANG and N.K. PAVLIS (1991) *The Ohio State 1991 Geopotential and Sea Surface Topography Harmonic Coefficient Models.* Number 410 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

REIGBER, C. (1969) *Zur Bestimmung des Gravitationsfeldes der Erde aus Satellitenbeobachtungen*, Volume 137 of the series C. Deutsche Geodätische Kommission, München.

REIGBER, C., P. SCHWINTZER and H. LÜHR (1999) The CHAMP geopotential mission. *Bolletino di Geofisica Teorica ed Applicata*, 40:285–289.

REUBELT, T., G. AUSTEN and E.W. GRAFAREND (2003) Harmonic analysis of the Earth's gravitational field by means of semi-continuous ephemerides of a low Earth orbiting GPS-tracked satellite. Case study: CHAMP. *Journal of Geodesy*, 77:257–278.

REUBELT, T., M. GÖTZELMANN and E.W. GRAFAREND (2006) Harmonic analysis of the Earth's gravitational field from kinematic CHAMP orbits based on numerically derived satellite accelerations. In: FLURY, J., R. RUMMEL, C. REIGBER, G. BOEDECKER M. ROTHACHER and U. SCHREIBER (Eds.), *Observation of the Earth System from Space.* Springer, 27–42.

REUTER, R. (1982) Über Integralformeln der Einheitssphäre und harmonische Splinefunktionen. *Veröffentlichungen des Geodätischen Instituts, RWTH Aachen*, 33.

RIESZ, F. (1907) Sur une espèce de géométrie analytiques des systèmes de fonctions sommables. *C. R. Acad. Sci. Paris*, 144:1409–1411.

ROWLANDS, D.D., S.B. LUTHKE, S.M. KLOSKO, F.G.R. LEMOINE, D.S. CHINN, J.J. MCCARTH, C.M. COX and O.B. ANDERSON (2005) Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophysical Research Letters*, 32:4310–+.

RUMMEL, R. (1986) Satellite Gradiometry. In: SÜNKEL, H. (Eds.), *Mathematical and numerical techniques in physical geodesy*, Lecture Notes in Earth Sciences 7. Springer, 220–312.

RUMMEL, R. (1992) *Fysische Geodesie I, Collegediktaat.* Technische Universiteit Delft.

RUMMEL, R and M VAN GELDEREN (1995) Meissl scheme - spectral characteristics of physical geodesy. *Manuscripta Geodaetica*, 20:379 – 385.

Rummel, R., K.P. Schwarz and M. Gerstl (1979) Least squares collocation and regularization. *Bulletin Geodesique*, 53:343–361.

Saff, E.B. and A.B.J. Kuijlaars (1997) Distributing many points on a sphere. *Math. Intelligencer*, 19:5–11.

Schmidt, H.F. (1981) Sampling function and finite element method representation of the gravity field. *Reviews of Geophysics and Space Physics*, 19(3):421–436.

Schmidt, M. (2001) *Grundprizipien der Wavelet-Analyse und Anwendungen in der Geodäsie.* Shaker Verlag, Aachen.

Schmidt, M., J. Kusche, J.P. van Loon, C.K. Shum, S.C. Han and O. Fabert (2005) Multiresolution representation of a regional geoid from satellite and terrestrial gravity data. In: Jekelis, C., L. Bastos and J. Fernandes (Eds.), *Gravity, Geoid and Space Missions.* Springer, 167–172.

Schmidt, M., M. Fengler, T. Mayer-Gürr, A. Eicker, J. Kusche, L. Sanchez and S.-C. Han (2006) Regional gravity modelling in terms of spherical base functions. *Journal of Geodesy*, 8:17–38.

Schmidt, M., S.-C. Han, J. Kusche, L. Sanchez and C.K. Shum (2006) Regional high-resolution spatiotemporal gravity modeling from GRACE data using spherical wavelets. *Geophysical Research Letters*, 33:8403–+.

Schmidt, R. (2007) Zur Bestimmung des cm-Geoids und dessen zeitlicher Variationen mit GRACE. *Scientific Technical Report STR07/04, GeoForschungsZentrum Potsdam.*

Schneider, M. (1968) A General Method of Orbit Determination. *Library Translation*, Volume 1279, Royal Aircraft Establishment, Ministry of Technology, Farnborough, England.

Schoenberg, I. (1964a) On interpolation by spline functions and its minimum norm properties. *Internat. Ser. Numer. Anal.*, 5:109–129.

Schoenberg, I. (1964b) Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. USA*, 52:947–950.

Schuh, W.D. (1996) Tailored Numerical Solution Strategies for the Global Determination of the Earth's Gravity Field. *Mitteilungen der geodätischen Institute der Technischen Universität Graz*, 81.

Schwarz, K.-P. (1971) *Numerische Untersuchungen zur Schwerefortsetzung.* Reihe C, Nr. 171. Deutsche Geodätische Kommission, München.

Seo, K.-W., C.R. Wilson, J. Chen and D.E. Waliser (2008) GRACE's spatial aliasing error. *Geophay. J. Int.*, 172:41–48.

Sharifi, M.A. and W. Keller (2005) GRACE gradiometer. In: Jekeli, C., L. Bastos and J. Fernandes (Eds.), *Gravity Geoid and Space Mission, GGSM 2004*, Volume 129 of the series International Association of Geodesy Symposia. Springer, 42–47.

Simons, F.J. and F.A. Dahlen (2006) Spherical Slepian functions and the polar gap in geodesy. *Geophysical Journal International*, 166:1039–1061.

Simons, F.J., F.A. Dahlen and M.A. Wieczorek (2006) Spatiospectral concentration on the sphere. *SIAM Review*, 48:504–536.

Slepian, D and H.O. Pollak (1960) Prolate spheroidal wave functions, Fourier analysis and uncertainty-I. *Bell Syst. Tech. J.*, 40:43–63.

Smith, S.W. (1997) *The Scientist and Engineer's Guide to Digital Signal Processing.* California Technical Pub., California.

SNEEUW, N. (1994) Global spherical harmonic analysis by least squares and numerical quadrature methods in historical perspective. *Geophys. J. Int.*, 118:707–716.

STANDISH, E. M. (1998) *JPL Planetary and Lunar Ephemerides DE405/LE405*. Jet Propulsion Labratory, Pasadena.

STROUD, A.H. and D. SECREST (1966) *Gaussian Quadrature Formulas*. Prentice –Hall, Englewood Cliffs, N.J.

STUHNE, G.R. and W.R. PELTIER (1999) New icosahedral grid–point discretizations of the shallow water equations on the sphere. *Journal of Computational Physics*, 148:23–58.

TAMMES, P.M-L. (1930) On the origin of number and arrangement of places of exit on the surface of pollen grains. *Recueil des Travaux Botanique Neerlandais*, 27:1–84.

TAPLEY, B., J. RIES, S. BETTADPUR, D. CHAMBERS, M. CHENG, F. CONDI, B. GUNTER, Z. KANG, P. NAGEL, R. PASTOR, T. PEKKER, S. POOLE and F. WANG (2005) GGM02 – An improved Earth gravity field model from GRACE. *Journal of Geodesy*, 79(8):467–478.

TAPLEY, B.D., S. BETTADPUR, M. WATKINS and CH. REIGBER (2004) The gravity recovery and climate experiment: mission overview and early results. *Geophys Res Lett*, 31.

THOMSON, J.J. (1904) On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Philosophical Magazine Series 6*, 7:237–265.

TIKHONOV, A.N. (1963) Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038.

TSCHERNING, C.C. (1972) Representation of Covariance Functions Related to the Anomalous Potential of the Earth using Reproducing Kernels. *The Danish Geodetic Institute Internal Report*, 3.

TSCHERNING, C.C. (1977) A note on the choice of norm when using collocation for the computation of approximations to the anomalous potential. *Bulletin Geodesique*, 51:137–147.

TSCHERNING, C.C. and R.H. RAPP (1974) *Closed covariance expressions for gravity anomalies, geoid undulations, and deflections of the vertical implied by anomaly degree variance models*. Number 208 in Reports of the Department of Geodetic Science. Ohio State University (OSU), Columbus, Ohio.

VAN DER CORPUT, J.G. (1935) Verteilungsfunktionen. *Proc. Ned. akad. v. Wet.*, 38:813–821.

VISSER, P., N. SNEEUW and C. GERLACH (2003) Energy integral method for gravity field determination from satellite orbit coordinates. *Journal of Geodesy*, 77(3–4):207–216.

VORONOI, G.F. (1908) Nouvelles applications des paramètres continus à la théorie de formes quadratiques. *J. Reine Angew. Math*, 134:198–287.

ŠVEHLA, D. and M. ROTHACHER (2001) Kinematic orbit determination of LEOs based on zero or double-difference algorithms using simulated and real SST GPS Data. In: ÁDÁM, J. and K.-P. SCHWARZ (Eds.), *Vistas for Geodesy in the new millenium*, Volume 125 of the series International Association of Geodesy Symposia. Springer, 322–328.

ŠVEHLA, D. and M. ROTHACHER (2003) Kinematic and reduced-dynamic precise orbit determination of low Earth orbiters. *Advances in Geosciences*, 1:47 – 56.

WAHBA, G. (1977) Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14(4):651–667.

WAHBA, G. (1981) Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comp.*, 1:5–16.

WAHBA, G. (1990) *Spline Models for Observational Data.* Regional conference series in applied mathematics. Society for industrial and applied mathematics, Philadelphia.

WEIGHTMANN, J.A. (1965) Gravity, geodesy and artificial satellites, a unified analytical approach. In: *Presented Paper: Symposium on the Use of Artificial Satellites for Geodesy.* Athen.

WHITE, D., A.J. KIMERLING and W.S. OVERTON (1992) Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems*, 19(1):5–22.

XU, P. (1992) The value of minimum norm estimation of geopotential fields. *Geophysical Journal International*, 111:170–178.

XU, P. and R. RUMMEL (1994) A simulation study of smoothness methods in recovery of regional gravity fields. *Geophysical Journal International*, 117:472–486.

XU, P. YUNZHONG, S., Y. FUKUDA and Y. LIU (2006) Variance component estimation in linear inverse ill-posed models. *Geophysical Journal International*, 80(2):69–81.